STATISTICAL ANALYSIS PLAN

Version: 2.0

Date: 26-Oct-2023

Protocol Number:	CPI 0610-04
Study Name:	MANIFEST-2
Study Title:	A Phase 3, Randomized, Double-blind, Active-Control Study of Pelabresib (CPI-0610) and Ruxolitinib vs. Placebo and Ruxolitinib in JAKi Treatment Naive MF Patients
Short Title:	Phase 3 Study of Pelabresib (CPI-0610) in MF
Compound:	Pelabresib (CPI-0610)
Phase:	3
IND Number:	147351
EudraCT Number:	2020-001989-10
Study Sponsor:	Constellation Pharmaceuticals, Inc. *a fully owned subsidiary of* *MorphoSys U.S. Inc.* 470 Atlantic Avenue, Suite 1401 Boston, MA, 02210 USA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

REVISION HISTORY

Version	Date	Primary Rationale for Revision
v1.0	Dec 15, 2022	Not applicable
v1.1	Jun 15, 2023	Correct exposure analysis
v1.2	July 15, 2023	Implementing Protocol Am5 (v 6.0)
v1.3	October 04, 2023	Incorporate comments from FDA type C meeting dated 19SEP2023
v2.0	October 26, 2023	Ready for signature

ABBREVIATIONS

AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine aminotransferase
AML	Acute Myelogenous Leukemia
aPTT	Activated Partial Thromboplastin Time
ARDS	Acute Respiratory Distress Syndrome
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AUC	Area Under the concentration-time Curve
BID	Twice daily
BET	Bromodomain and Extra-Terminal Motif
BUN	Blood Urea Nitrogen
CDF	Cumulative distribution function
CI	Clinical Improvement
CHR	Complete Hematological Response
C_{max}	Concentration maximum
CR	Complete Response/Remission
CRO	Contract Research Organization
CRP	C-reactive Protein
CT	Computerised Tomography
CTCAE	Common Terminology Criteria for Adverse Events
C_{trough}	Concentration trough
CSR	Clinical Study Report
DIC	Disseminated Intravascular Coagulation
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
ELN	European Leukemia Net
eCRF	Electronic Case Report Form
EMA	European Medicines Agency

EOS	End of Study
EOT	End of Treatment
EPO	Erythropoietin
ET	Essential Thrombocythemia
FDA	Food and Drugs Administration
GCP	Good Clinical Practice
HDL	High Density Lipoprotein
HU	Hydroxyurea
Hgb	Hemoglobin
ICF	Informed Consent Form
ICH	International Council for Harmonisation
INR	International Normalized Ratio
ITT	Intention-to-Treat
IWG	International Working Group
K-M	Kaplan-Meier
LDH	Lactate Dehydrogenase
LDL	Low Density Lipoprotein
MDS	Myelodysplastic Syndrome
MedDRA	Medical Dictionary for Regulatory Activities
MF	Myelofibrosis
MFSAF	Myelofibrosis Symptom Assessment Form
mg	Milligram
mL	Milliliter
MPN	Myeloproliferative Neoplasms
MPN-SAF	Myeloproliferative Neoplasms Symptom Assessment Form
mRNA	Messenger Ribonucleic Acid
MRI	Magnetic Resonance Imaging
MRT	Myeloproliferative Neoplasms Research and Treatment
NCI	National Cancer Institute
ng	Nanogram
OHR	Overall Hematological Response
PD	Protocol Deviation

PE	Physical Examination
PGIC	Patient Global Impression of Change
PK	Pharmacokinetics
PO	Per os; Orally
PR	Partial Response/Remission
PRO	Patient Reported Outcome
PT	Preferred Term
RBC	Red Blood Cell
RNA	Ribonucleic Acid
SAP	Statistical Analysis Plan
SAE	Serious adverse event
SAF	Safety Set
SD	Stable Disease
StD	Standard Deviation
SoC	System Organ Class
SOP	Standard Operating Procedure
SVR	Spleen Volume Reduction
TD	Transfusion Dependent
TE	Thromboembolic
TEAE	Treatment Emergent Adverse Event
TI	Transfusion Independent
T_{max}	Time to Maximum Concentration
TSS	Total Symptom Score
ULN	Upper Limit of Normal
WBC	White Blood Cell
WHO	World Health Organisation

1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to provide a detailed and comprehensive description of the planned statistical analyses for the study with protocol number CPI0610-04, a Phase 3, global, multicenter, randomized, double-blind, active-controlled study of pelabresib + ruxolitinib vs placebo + ruxolitinib in Janus kinase (JAK) inhibitor treatment naïve patients with primary myelofibrosis (PMF), post-polycythemia vera (PV) myelofibrosis (PPV-MF) or post-essential thrombocythemia (ET) myelofibrosis (PET-MF).

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials. All work planned and reported from this SAP will follow internationally accepted guidelines for statistical practice, published by the American Statistical Association and the Royal Statistical Society.

Populations for analysis, data handling rules, statistical methods, and formats for data presentation are provided.

The planned analyses identified in this SAP will be included in clinical study reports (CSR) and/or in relevant summary report documents (e.g., regulatory submissions, abstract submissions, or future manuscripts). Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data and will not require updating the final SAP. Any post-hoc, or unplanned, exploratory analysis performed will be clearly identified as such and described in the final Clinical Study Report (CSR).

The following documents were reviewed in preparation of this SAP:

Study CPI 0610-04 Protocol, Version 4, dated 15- Nov-2021

Patient Case Report Forms (eCRF), dated 31-Aug-2021

ICH Guidance on Statistical Principles for Clinical Trials (E9)

The reader of this SAP is encouraged to also read the aforementioned documents, for details on the design and planned conduct of this study. Any amendments to the protocol, which do not affect the statistical analyses, will not necessitate an update to this document.

2 STUDY OBJECTIVES, ENDPOINTS, AND ESTIMANDS

The table below reports the objectives and endpoints of the Phase 3 study, with indication on the primary, secondary, and exploratory nature of the endpoint.

2.1 Primary Objective and Endpoint

Primary Objective	Endpoint
To determine the efficacy of pelabresib + ruxolitinib compared with placebo + ruxolitinib	Splenic response as assessed by central radiology read at Week 24

2.2 Key Secondary Objective and Endpoint

Key Secondary Objective	Endpoint
To determine the effect of pelabresib + ruxolitinib on the absolute change in TSS at Week 24 vs Baseline compared with placebo + ruxolitinib	Absolute change in TSS at Week 24 compared to Baseline
To determine the efficacy of pelabresib + ruxolitinib compared with placebo + ruxolitinib	TSS50 response at Week 24

2.3 Secondary Objectives and Endpoints

Secondary Objectives	Endpoints
To determine the effect of pelabresib + ruxolitinib on the percent change in total symptom score (TSS) at Week 24 vs Baseline compared with placebo + ruxolitinib	Percent change in TSS at Week 24 compared to Baseline
To characterize the effects of pelabresib + ruxolitinib compared with placebo + ruxolitinib on bone marrow fibrosis	Improvement in bone marrow fibrosis by at least 1 grade at Week 24 compared to Baseline, as assessed by central review
To determine the effect of pelabresib + ruxolitinib on the durability of splenic response compared with placebo + ruxolitinib	Splenic response as assessed by central review at Week 48
To determine the effect of pelabresib + ruxolitinib on the durability of TSS response compared with placebo + ruxolitinib	TSS50 response at Week 48
To determine the effect of pelabresib + ruxolitinib on the durability of absolute change in TSS at Week 48 vs Baseline compared with placebo + ruxolitinib	Absolute change in TSS at Week 48 compared to baseline
To determine the effect of pelabresib + ruxolitinib compared with placebo + ruxolitinib on the rate of red blood cell (RBC) transfusion over the first 24 weeks of treatment	Rate of RBC transfusion over the first 24 weeks of treatment
To determine the effect of pelabresib + ruxolitinib compared with placebo + ruxolitinib on the conversion from RBC transfusion dependence to independence	Conversion from RBC transfusion dependence to independence
To evaluate the Patient Global Impression of Change (PGIC) at Week 24	Category change of PGIC at Week 24 compared to Baseline

Secondary Objectives	Endpoints
To evaluate progression-free survival (PFS) in patients treated with pelabresib + ruxolitinib compared with placebo + ruxolitinib	PFS where progression is defined as per International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) response criteria
To evaluate overall survival (OS) in patients treated with pelabresib + ruxolitinib compared with placebo + ruxolitinib	OS
To evaluate acute myeloid leukemia (AML) conversion in patients treated with pelabresib + ruxolitinib compared with placebo + ruxolitinib	Proportion of patients with transformation to blast phase (AML)
To determine the safety and tolerability of pelabresib + ruxolitinib compared with placebo + ruxolitinib	Adverse events (AEs) of all grades and Serious adverse events (SAEs)
To characterize the pharmacokinetics (PK) of pelabresib	Population PK assessment including determination of exposure metrics and secondary parameters (i.e., $AUC_{0\text{-}t}$, T_{max}, C_{max}, $T_{1/2}$, Vd/F, CL/F)
To characterize the effects, if any, of pelabresib on the PK of ruxolitinib	Descriptive assessment of ruxolitinib plasma concentrations in the presence or absence of pelabresib
To determine the effect of pelabresib + ruxolitinib on the duration of splenic response compared with placebo + ruxolitinib	Duration of the splenic response
To determine the effect of pelabresib + ruxolitinib on the modified TSS compared with placebo + ruxolitinib	Modified TSS response at Week 24 (TSS score without the fatigue sub-domain)
To determine the effect of pelabresib + ruxolitinib on the duration of TSS response compared with placebo + ruxolitinib	Duration of the TSS response

2.4 Exploratory Objectives and Endpoints

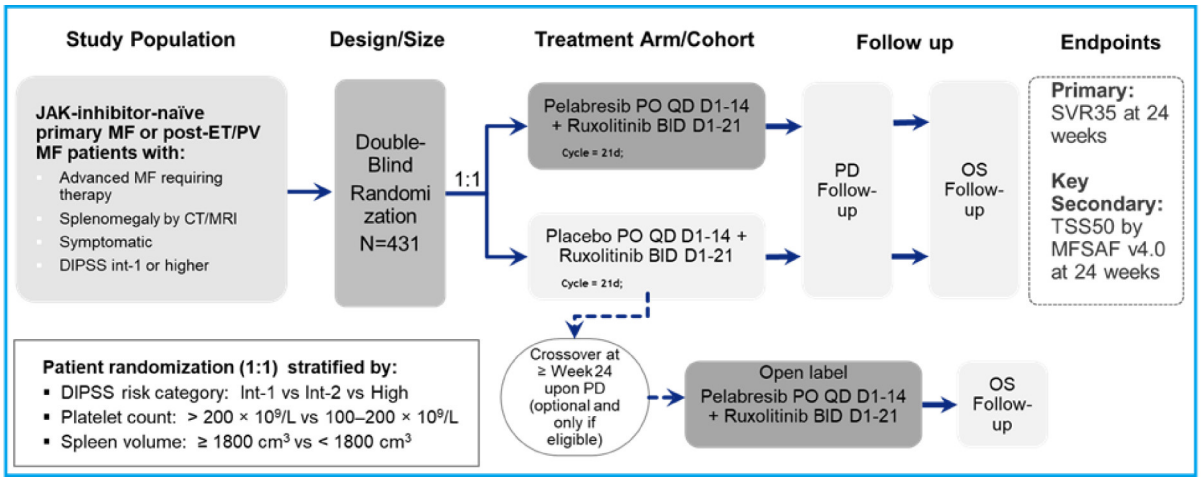
Exploratory Objectives	Endpoints
To determine the efficacy of pelabresib + ruxolitinib compared with placebo + ruxolitinib on the percent change in splenic volume at Week 24	Percent change in splenic volume at Week 24 as per central read
To determine the effect of pelabresib + ruxolitinib compared with placebo + ruxolitinib on RBC transfusion dependence rate	RBC transfusion dependence at Week 24
To determine the effect of pelabresib + ruxolitinib compared with placebo + ruxolitinib on hemoglobin response	Hemoglobin response
[redacted]	

Exploratory Objectives	**Endpoints**
[redacted]	
To characterize the PD effects of pelabresib + ruxolitinib compared with placebo + ruxolitinib in the blood	• Post-treatment changes from Baseline in circulating concentrations of cytokines • Post-treatment changes from Baseline in the ratio of mutant to wild type JAK2, [redacted]
[redacted]	
To determine the effect of pelabresib + ruxolitinib on the time to splenic response compared with placebo + ruxolitinib	Time to first SVR35 response
[redacted]	

3 Overall Design

This is a Phase 3, global, multicenter, randomized, double-blind, active-controlled study of pelabresib + ruxolitinib vs placebo + ruxolitinib in JAKi-treatment naïve patients with PMF, PPV-MF, or PET-MF. The study design schematic is shown in Figure 1.

Figure 1: Study Design Schematic



Note: PD=Progressive Disease

The 28-day screening period begins the day the patient signs the Informed Consent Form (ICF). Patients deemed eligible for enrollment will be randomized in a 1:1 ratio to 1 of 2 treatment groups: 1) pelabresib + ruxolitinib (experimental group), or 2) placebo + ruxolitinib (control group). Dosing should begin as soon as possible following randomization and both randomization and dosing (C1D1) must occur within 28-days from the beginning of screening. Patients who have not been randomized during the screening period can be rescreened if they did not meet the criteria for participation in this study (screen failure).

Patients who are considered screen failures only based on Inclusion Criterion #7 (Have at least 2 symptoms with an average score ≥ 3 over the 7-day period prior to randomization or an average total score of ≥ 10 over the 7-day period prior to randomization using the MFSAF v4.0) may be rescreened only after a 6-week period. Patients who have been randomized, but not dosed within 28-days since the start of Screening cannot be rescreened and will be terminated from the study.

Patients will be assessed for the primary and key secondary endpoints, splenic and TSS response, respectively, at Week 24. After 24 weeks, patients in the control group who have progressive disease due to splenomegaly (i.e., defined as enlargement of spleen volume by MRI or CT scan of ≥25% compared to the baseline value, as confirmed by the central review; not due to leukemic transformation or an increase in peripheral blood blast percentage of ≥20% that persists for at least 1 week) may have the option to cross over to receive the experimental treatment of ruxolitinib + pelabresib in an open label fashion. Data collected as part of the crossover period will be analyzed descriptively.

Double-blind treatment (pelabresib or matching placebo) will be administered QD for 14 consecutive days followed by a 7-day break, which is considered 1 cycle of treatment (1 cycle = 21 days). Ruxolitinib will be administered BID for all 21 days within each cycle. All patients who have progressive disease at any time will be required to discontinue the double-blind study treatment.

Approximately 400 patients (200 per treatment arm) will be enrolled in the study.

An independent DSMB will be used to evaluate safety on an ongoing basis throughout the study.

3.1 Schedule of Activities

Table 1: Schedule of Activities

Procedure	**Screening Period**	**Treatment Period (including crossover)**						**Follow-Up Period[b]**		
	Screening (within 28 days of C1D1)	**C1 or initiation of crossover treatment[a]**		**C2-8**	**C9 and all odd cycles after[f]**	**C10 and all even cycles after**	**EOT (within 7 days after last dose of pelabresib/placebo)**	**Safety F/U (30 days [±3d] after last dose of pelabresib/placebo)**	**Progressive disease F/U (every 3 months [±14d] after last MRI/ CT scan)**	**Survival F/U (every 3 months [±14d] after PD)**
		D1	**D14 (±1d)**	**D1 (±3d)**	**D1 (±3d)**	**D1 (±3d)**				
Informed consent	X									
Inclusion and exclusion criteria	X									
Demographics	X									
Medical history	X									
Full PE and VS	X	X								
Targeted PE and VS			X	X	X		X			
Height	X									
Weight	X	X	X	X	X		X			
MF risk category	X									
Transfusion history	X	X		X	X	X	X		X	
ECOG status	X	X		X			X			

Procedure	**Screening Period**	**Treatment Period (including crossover)**						**Follow-Up Period[b]**		
	Screening (within 28 days of C1D1)	**C1 or initiation of crossover treatment[a]**		**C2-8**	**C9 and all odd cycles after[f]**	**C10 and all even cycles after**	**EOT (within 7 days after last dose of pelabresib/placebo)**	**Safety F/U (30 days [±3d] after last dose of pelabresib/placebo)**	**Progressive disease F/U (every 3 months [±14d] after last MRI/ CT scan)**	**Survival F/U (every 3 months [±14d] after PD)**
		D1	**D14 (±1d)**	**D1 (±3d)**	**D1 (±3d)**	**D1 (±3d)**				
IWG-MRT response evaluation	X	X			Cycle 9 only					
12-lead ECG[c]	X	X		X (C2, C3, C5, C7 only)	X		X			
Serum or urine pregnancy test (WOCBP only)[d]	X	X		X	X	X	X and monthly thereafter until 184 days after the last dose of study drug			
Hematology	X	X (within 72 hrs of C1D1)[e]	X	X	X		X		X	
Clinical chemistry	X	X (within 72 hrs of C1D1)[e]	X	X	X		X			
Coagulation parameters	X	C3 D1, and every other cycle thereafter					X			
Iron studies	X	C5 D1, C9 D1, and every 4 cycles thereafter								

Procedure	**Screening Period**	**Treatment Period (including crossover)**						**Follow-Up Period[b]**		
	Screening (within 28 days of C1D1)	**C1 or initiation of crossover treatment[a]**		**C2-8**	**C9 and all odd cycles after[f]**	**C10 and all even cycles after**	**EOT (within 7 days after last dose of pelabresib/placebo)**	**Safety F/U (30 days [±3d] after last dose of pelabresib/placebo)**	**Progressive disease F/U (every 3 months [±14d] after last MRI/ CT scan)**	**Survival F/U (every 3 months [±14d] after PD)**
		D1	**D14 (±1d)**	**D1 (±3d)**	**D1 (±3d)**	**D1 (±3d)**				
HbA1c and serum lipid panel	X	Every 6 months after screening (follow the local approved ruxolitinib labeling for the frequency of lipid monitoring if different)					X			
TB testing	X (per local regulations)									
Assess hepatitis B/C, HIV, and COVID-19 risk and test if indicated	X									
MRI (or CT) scan	X	Every 12 weeks from C1 D1 (window ±14 days) irrespective of any delay in cycle due to drug hold until PD or initiation of another anti-cancer therapy MRI is preferred method. Method of assessment used at screening should remain consistent throughout study. Repeated at EOT only if PD has not been previously documented (or in the absence of PD, if imaging has not been performed within the previous 6 weeks).								
Bone marrow biopsy	X (only if not done within 12 weeks of C1 D1)	Every 24 weeks (window ±14 days) from C1 D1 to Week 72, then every 48 weeks thereafter (window ±28 days), irrespective of any delay in cycle due to drug hold, and at EOT (EOT sample does not need to be collected if performed within previous 12 weeks), until PD or initiation of another anti-cancer therapy								
MFSAF v4.0	X (daily for at least 5 of 7 days before randomization)	Once daily (completed around the same time each day) until 12 weeks after EOT								
PGIC		Once weekly (completed on the same day each week) until 12 weeks after EOT								

Procedure	**Screening Period**	**Treatment Period (including crossover)**						**Follow-Up Period[b]**		
	Screening (within 28 days of C1D1)	**C1 or initiation of crossover treatment[a]**		**C2-8**	**C9 and all odd cycles after[f]**	**C10 and all even cycles after**	**EOT (within 7 days after last dose of pelabresib/placebo)**	**Safety F/U (30 days [±3d] after last dose of pelabresib/placebo)**	**Progressive disease F/U (every 3 months [±14d] after last MRI/ CT scan)**	**Survival F/U (every 3 months [±14d] after PD)**
		D1	**D14 (±1d)**	**D1 (±3d)**	**D1 (±3d)**	**D1 (±3d)**				
EQ-5D	X (once weekly)	Once weekly (completed on the same day each week) until 12 weeks after EOT								
PK		X	X	C3D1 and C7D1	X (C9 only)					
Biomarker sampling		X	X		X (C9, C17 only)					
Randomization	X									
Pelabresib or placebo administration		QD for first 14 days of each cycle If dose increases on even cycle, odd cycle activities should be followed.								
Ruxolitinib administration		BID for all 21 days of each cycle If dose increases on even cycle, odd cycle activities should be followed.								
AE review		Throughout								
Concomitant medication review		Throughout								
Subsequent anti-MF therapy review								Throughout		

AE = adverse event; BID = twice daily; C = cycle; CT = computerized tomography; D = day; ECG = electrocardiogram; ECOG = Eastern Cooperative Oncology Group; EOT = end of treatment; F/U = follow-up; MFSAF = Myelofibrosis Symptom Assessment Form; MRI = magnetic resonance imaging; PE = physical exam; PGIC = Patient Global Impression of Change; PK = pharmacokinetic; QD = once daily; QTcF = Fridericia-corrected QT interval; TB = tuberculosis; VS = vital signs; WOCBP = women of childbearing potential.

Note: Schedule of activities are based on protocol amendment v5.0

Note: Laboratory assessments and other safety assessments must be performed predose when a window is specified.

Note: Assessments performed at unscheduled visits due to safety reasons are at the discretion of the Investigator.

a Patients who cross over should be re-consented at the initiation of crossover treatment.

b Follow-up visits may be conducted via telephone when clinic visits are not required. Patients who discontinue treatment for reasons other than documented disease progression should have follow-up visits every 12 weeks to document response by imaging, and transfusion requirements until initiation of another anti-cancer therapy, disease progression, death, or the end of the study, whichever comes first (progressive disease F/U period). Continuing ruxolitinib monotherapy as standard of care treatment after the discontinuation of treatment on study is not considered the initiation of another anti-cancer medication/therapy. Patients who discontinue study treatment for documented disease progression or start another anti-cancer medication/therapy should have a follow-up visit/phone call every 12 weeks to document overall survival (Survival F/U). Information on subsequent anti-MF medication will be collected in both progressive disease F/U and Survival F/U periods.

c QTcF (Fridericia) is to be used at screening to evaluate eligibility and throughout the study to maintain consistency.

d For WOCBP, pregnancy testing (serum preferred) at Screening and at C1D1 before study drug administration; thereafter pregnancy testing can be either highly sensitive urine or serum testing at D1 of every Cycle before study drug administration, at EOT and thereafter at monthly intervals during 184 days after last dose of study drug; urine pregnancy testing may be conducted at home when clinic visits are not required. The Investigator must be informed immediately about the results of home pregnancy tests, and the results are to be recorded by the Investigator.

e Laboratory assessments for C1D1 do not need to be repeated, and screening values may be used, if performed within 72 hours of C1D1 assessments.

f After C9D1 patients may come in every odd cycle if they meet criteria.

3.2 Randomization and Blinding

This study has a double-blind design in which patients and investigators are blinded to study drug; pelabresib and placebo will be packaged identically.

All patients will be randomly assigned to either treatment group in a 1:1 ratio using a centralized IVRS/IWRS. Before the study is initiated, operational logistics for the IVRS will be provided to each site in an IRT manual. Note: dosing should begin as soon as possible following randomization, and no later than 28 days from start of the screening period.

The IVRS/IWRS will be programmed with blind-breaking instructions. In case of an emergency, the Investigator has the responsibility for determining if unblinding a patient's intervention assignment is warranted. Patient safety must always be the first consideration in making such a determination. If the Investigator decides that unblinding is warranted, the Investigator should make every effort to contact the Sponsor prior to unblinding a patient's intervention assignment unless this could delay emergency treatment of the patient. If a patient's intervention assignment is unblinded, the Sponsor must be notified within 24 hours after breaking the blind. The date the blind was broken must be recorded in the source documentation and IVRS/IWRS, as applicable.

Each patient in the study is identified by a unique patient number. The procedures for patient numbering and coordination between the study sites will be provided in a separate document prior to study start. The unique patient number is a combination of his/her study center number and a second number reflecting the sequence of patient enrollment. The study center number is assigned by the Sponsor to each investigative site. Upon signing the ICF, the patient is assigned a patient number. Once assigned to a patient, a patient number will not be reused.

To ensure balance between treatment arms, patient randomization will be stratified by the following factors:

- DIPSS risk category: Intermediate-1 vs. Intermediate-2 vs. High
- Platelet count: $> 200 \times 10^9$/L vs. $100 – 200 \times 10^9$/L
- Spleen volume: ≥ 1800 cm^3 vs. < 1800 cm^3

3.3 Planned Analyses

3.3.1 Periodic Safety Analyses (Safety DSMB Meetings)

An independent Data and Safety Monitoring Board (DSMB) is established to act in an expert, advisory capacity for periodic assessment of the data to monitor participant safety and to ensure the validity and scientific merit of the trial.

The analyses conducted in the context of the DSMB meetings and the DSMB schedules are specified in the DSMB SAP.

3.3.2 Primary Efficacy Analysis

3.3.2.1 Purpose and Timing

The purpose of the Primary Analysis is to determine whether the treatment with pelabresib + ruxolitinib is (statistically) significantly superior to ruxolitinib monotherapy in terms of efficacy. The treatments will also be compared in terms of safety in a descriptive manner. All results of the analyses specified in this SAP will be reported in a Clinical Study Report (CSR).

The Primary Analysis will take place after all randomized patients have either completed the 24 weeks efficacy assessments or been prematurely discontinued. By this time, it is expected that at least 50% of the randomized patients will have completed their 36 weeks efficacy assessments.

3.3.2.2 Analyses to be Performed

All analyses specified in this SAP will be performed except for time-to-event analyses with limited event sizes.

3.3.2.3 Data Cleaning and Data Cut-off

The study CRO and the sponsor will be closely monitoring patient disposition and the scheduled Week 24 assessments for each patient (based on regular database extractions), and the data cut-off date will be the date when the last patient's Week 24 (Cycle 9 Day 1 visit) data have been captured in the database.

All data required for the Primary Analysis will be cleaned, reconciled, source data verified as applicable and signed by the Investigator. Details are documented in the data management plan (DMP).

Any data collected beyond the cut-off date will not be included in the analysis. 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. For example, if the cut-off date is 15 June 2023, then an AE starting on 13 June 2023 will be reported, whereas an AE with start date on 17 June 2023 will not be reported.

All events with an event start date either before or on the cut-off date and an event end date after the cut-off date will be reported in listings as "ongoing" (the end date is missing in listings). The same rule applies to events starting either before or on the cut-off date and not having documented end date.

If it is required to impute an end date, the missing end date is replaced by the cut-off date and is flagged in the listings.

3.3.2.4 Extent of Blinding

Until the Primary Analysis takes place, patients, investigators, and both the sponsor and CRO study team will remain blinded to patient allocation to study arms, except for those patients who were enrolled into the crossover treatment period at or after Week 24 and therefore had to be unblinded.

3.3.3 Final Analysis

3.3.3.1 Purpose and Timing

The Final Analysis will be performed at study end. The final analysis is going to happen after all patients have completed or prematurely discontinued the study treatment period or closed due to other reasons. The end of study is expected to occur approximately 5 years after the first patient was enrolled. All endpoints, including OS, will be analyzed in the final analysis.

3.3.3.2 Analyses to be Performed

All analyses specified in this SAP will be conducted.

3.3.3.3 Data Cleaning

Data will be cleaned, reconciled, source data verified, signed, and locked. Please refer to the DMP for further details.

3.4 Sample Size Determination

For sample size calculation, it is assumed that 62% and 29% patients will achieve splenic response for the primary efficacy endpoint, in the experimental group and the control group, respectively. It is assumed that 57% and 42.2% patients will achieve a TSS response at Week 24 for the key secondary efficacy endpoint, in

the experimental group and the control group, respectively (Mesa et al. 2017; Mascarenhas et al. 2019). A 2-sided 95% confidence interval (CI) of the difference in the proportions of patients with splenic response will be calculated. A sample size of 200 patients in each treatment group (400 patients in total) will provide > 99% power for testing the primary endpoint and 81% power for testing the key secondary endpoint TSS50, using the 2-group continuity corrected χ^2 test with a 5% two-sided significance level and accounting for 2% non-evaluable patients. This sample size will also provide more than 90% power to reject the null hypothesis of equal means when mean absolute TSS change from baseline difference between two arms is 4 with a standard deviation for both groups of 12 using the two-sample t-test with a 5% two-sided significance level. The assumption of absolute change difference of 4 points is based on the median percentage change difference observed from Mesa et al. 2017 and Mascarenhas et al. 2019.

4 DEFINITIONS

4.1 General definitions

4.1.1 Treatment Arms

- Arm 1 - experimental treatment group: Patients treated with pelabresib + ruxolitinib
- Arm 2 - control group: Patients treated with placebo + ruxolitinib

4.1.2 Treatment Period

- Double-blind treatment period – randomized treatment period.
- Crossover treatment period – open label pelabresib + ruxolitinib treatment. This is optional for patients who were randomized to the control group, and not applicable for patients who were randomized to the experimental treatment group.

4.1.3 Study Drug

Study drug refers to pelabresib, pelabresib matching placebo, and ruxolitinib components.

4.1.4 Treatment Cycle

A complete treatment cycle is defined as 21 calendar days during which pelabresib/placebo is administered for 14 consecutive days followed by a 7-day break (1 cycle = 21 days, i.e., 14 days of pelabresib/placebo + 7-day break). For each cycle, ruxolitinib is administered on a continuous basis.

4.1.5 Dates of Study Drug Administration

Date of first administration of study drug, which is also named as 'Start date of study drug', refers to the date of first non-zero dose of administration of any study drug.

Date of last administration of study drug, which is also named as 'End date of study drug', refers to the date of last non-zero dose of administration of any study drug. For reporting purposes, the date of last dose of pelabresib/pelabresib matching placebo will represent the end of combination treatment.

4.1.6 Reference Start Date and Study Day

For this randomized study, unless otherwise specified, the **reference start date for all safety assessments** (e.g., adverse event onset, laboratory/ vital sign measurement, study drug administration, prior and concomitant medication, medical history etc.), demographic and baseline assessment, pharmacokinetics analysis and biomarker analysis will be the date of first administration of any study drug.

Unless otherwise specified, the **reference start date for efficacy assessments** and duration of overall survival during follow-up period will be the randomization date.

The reference start date is designated as Study Day 1. Study Day –1 is the day that precedes Study Day 1. Study Day 0 is not defined.

The **study day** describes the day of the event or assessment date, relative to the reference start date.
The study day will be calculated as:

- If the event/assessment is on or after the reference start date, study day = [Date of the event/assessment - reference start date + 1]
- If the event/assessment precedes the reference start date, study day = [Date of the event/assessment - reference start date]

The study day will be displayed in the data listings.

4.1.7 Time Unit

Unless specifically mentioned, month will be used as the time unit for analysis. A month-length is 30.4375 days (365.25/12). If duration is to be reported in months, duration in days is divided by 30.4375. If duration is to be reported in years, duration in days will be divided by 365.25.

4.1.8 Baseline

For efficacy evaluations, the last non-missing assessment, including unscheduled assessments on or before the date of randomization is taken as "baseline" value or "baseline" assessment.

For safety evaluations, unless otherwise stated, the last non-missing assessment, including unscheduled assessments on or before the date of start of study treatment is taken as "baseline" value or "baseline" assessment. Assessments that are repeated on the first day of treatment post-dose will not be considered as baseline values.

The crossover baseline is defined as the last value for the assessment prior to taking the 1st dose of pelabresib in the crossover period.

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

Change from Baseline Calculation

Absolute change from baseline will be calculated as

[visit value – baseline value]

Percentage change from baseline will be calculated as

$$[\frac{\text{visit value}-\text{baseline value}}{\text{baseline value}} \times 100\ ].$$

4.2 Analysis Set

4.2.1 Screened Set

The Screened Set consists of all patients who signed the ICF and completed the informed consent eCRF page.

4.2.2 Intent-to-Treat (ITT) Analysis Set

The ITT set will include all randomized patients, for whom a randomization number has been assigned. This is the primary population for all efficacy endpoints. All analyses using this population will be based on the treatment assigned by the Interactive Response Technology (IRT) system.

4.2.3 Modified Intent-to-Treat (mITT) Analysis Set

The modified Intent-to-Treat (mITT) set will include all randomized patients who have received at least 1 dose of any study drug(s). This is the population for some sensitivity analyses on efficacy endpoints. All analyses using this population will be based on the treatment assigned by the IRT system.

4.2.4 Safety Analysis Set

A subset of the ITT population that includes all randomized patients who were administered at least one dose of study drug. This population will be used for the safety analyses. All analyses using this population will be based on the actual treatment received.

4.2.5 Crossover Set

This population will include all patients from the control arm who crossed over to receive open label pelabresib + ruxolitinib treatment and have received at least one dose of pelabresib.

4.2.6 Per-Protocol (PP) Set

The per-protocol (PP) set will be a subset of the ITT population that includes patients who have received adequate exposure to study therapy, i.e., a minimum of 4 cycles of treatment, and do not have any protocol deviations that could confound the interpretation of the primary analyses conducted on the ITT population. All analyses using this population will be based on the actual treatment received.

All protocol deviations will be specified in the "Protocol deviation specification form" and protocol deviations leading to exclusion from the PP will be decided in blinded fashion by the trial statistician, clinical trial lead, clinical development, and regulatory lead prior to data extraction for any analysis that would use PPS.

The protocol deviations that lead to exclusion from the PP Set may include but are not limited to the following:

- Taking wrong study treatment or receiving no study treatment
- Violations of any of the key inclusion/exclusion criteria
- Received prohibited concomitant medications while on treatment that may affect spleen size or TSS
- Substantial non-compliance with study treatment
- Unable to perform Week 24 spleen volume assessment or collect Week 24 TSS scores analyses due to reasons other than lack of efficacy or treatment related AEs
- Deviation that leads to significant bias of the treatment effect

However, deviations that might be affected by the actual treatment should not be used as exclusion criteria.

All protocol deviations will be specified in the "Protocol deviation specification form" and protocol deviations leading to exclusion from the PP set will be tabulated.

4.2.7 PK Analysis Set

The PK analysis set will include all patients in the Safety Set who received placebo/pelabresib + ruxolitinib treatment and have plasma samples drawn and tested for study drug concentrations.

4.2.8 Biomarker Analysis Set

The Biomarker Analysis set will include all patients in the Safety Set who received any treatment component and have evaluable biomarker data collected.

5 MEDICAL CODING

Coding for adverse events, medical history, prior systemic therapy for cancer, concomitant medication, and concomitant procedures will be performed using the Medical Dictionary for Regulatory Activities (MedDRA) and WHO Drug dictionary (WHO-DD).

Conditions and procedures will be coded using MedDRA version 26.0. Entries will be grouped by System Organ Class and Preferred Term.

Prior, concomitant, new anti-cancer medications and pre-medications will be coded using the WHO-DD version WHODD B3 Global March 2021 or later. Entries will be grouped by Anatomical Therapeutic Chemical (ATC) Level 4 class and preferred name.

Toxicity grade of adverse events and medical history are captured directly from eCRF. Toxicity grade of local laboratory data will be coded according to National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) V5.0.

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is 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 twice as bad as a grade 1).

If CTCAE grading does not exist for an adverse event, grades 1 – 4 corresponding to the severity of mild, moderate, severe, and life-threatening will be used.

6 STATISTICAL ANALYSIS

6.1 General Considerations

All hypothesis testing and confidence intervals (CIs) will be 2-sided. An alpha of 0.05 will be used for all statistical testing, unless otherwise stated. All p-values reported will be 2-tailed and rounded to 3 decimal places prior to assessment of statistical significance.

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. CIs will be presented using the same number of decimal places as the parameter estimate.

Where applicable, variables will be summarized descriptively by study visit. For the categorical variables, the counts and proportions of each possible value will be tabulated by the treatment group. The denominator for the proportion will be based on the number of patients in each category. In case of subcategories, the relative frequencies will be calculated based on the patients in the respective subcategory, in which case a footnote will be added explaining the different denominators.

A windowing convention will be used to determine the analysis values for a given study visit for observed data analyses.

6.1.1 Handling of Treatment Misallocations

For summary or analysis of Safety Set, the actual treatment is used, otherwise, the randomized treatment is used.

6.1.2 Analysis Approach for Continuous Variables

The values of continuous variables collected at baseline and subsequent study visits will be summarized, and the number of patients with non-missing values, mean, median, interquartile ranges (Q1, Q3, if applicable), SD, minimum, and maximum values will be tabulated.

6.1.3 Analysis Approach for Binary Variables

Certain binary outcome variables (i.e., response variables) will be analysed by stratified Cochran-Mantel-Haenszel (CMH) tests. Treatment groups will be compared in a pairwise fashion with respect to the proportion of subjects with “response” using the CMH test statistics. The corresponding p-values will be based on the CMH statistics which follows a Chi-square distribution with one degree of freedom.

Response variables will be coded 1 for response and 0 for non-response.

6.1.4 Analysis Approach for Time-to-Event Variables

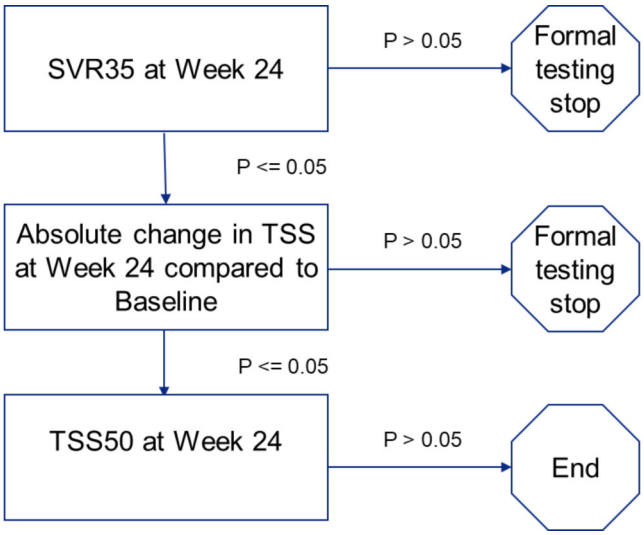
Time-to-event endpoints will be summarized using Kaplan-Meier estimates. 95% confidence intervals for the estimated 25%, 50%, and 75% times will be presented.

6.1.5 Statistical Hypotheses and Decision Rules

6.1.5.1 Statistical HypothesesStatistical Hypotheses

The hierarchical testing strategy for MANIFEST-2 is presented in Figure 2.

Figure 2: Sequential Testing of Pre-specified Endpoints



Primary Endpoint – SVR35

The null hypothesis and alternative hypothesis for analysis of the primary endpoint (H1) are:

$$H1_0: P1_E - P1_C = 0$$

vs.

$$H1_1: P1_E - P1_C \neq 0$$

$P1_E$: proportion of responders in **experimental group** (pelabresib + ruxolitinib) who achieve ≥35% reduction from baseline spleen size (SVR35 response) at Week 24

$P1_C$: proportion of responders in **control group** (placebo + ruxolitinib) who achieve ≥35% reduction from baseline spleen size (SVR35 response) at Week 24

For the main analysis, Cochran-Mantel Haenszel test using the stratification factors as per randomization on ITT will be used to compare SVR35 response rates between two arms.

Key Secondary Endpoint – Absolute Change from Baseline in TSS

The null hypothesis and alternative hypothesis for analysis of the key secondary endpoint (H2) are:

$$H2_0: \mu_E - \mu_C = 0$$

vs.

$$H2_1: \mu_E - \mu_C \neq 0$$

μ_E: mean absolute change from baseline in TSS at Week 24 for **experimental group** (pelabresib + ruxolitinib)

μ_C: mean absolute change from baseline in TSS at Week 24 for **control group** (placebo + ruxolitinib)

For the main analysis, ANCOVA (Analysis of Covariance) model will be used to analyze the absolute change from baseline in TSS at Week 24. The dependent variable is change from baseline in TSS at Week 24, with treatment group, baseline DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), baseline platelet count (100-200 × 10^9/L vs. > 200 × 10^9/L) and baseline spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) as fixed effects and the baseline TSS as a covariate.

Key Secondary Endpoint – TSS50

The null hypothesis and alternative hypothesis for analysis of the key secondary endpoint (H3) are:

$$H3_{0:}\ P2_E - P2_C = 0$$

vs.

$$H3_{1:}\ P2_E - P2_C \neq 0$$

$P2_E$: proportion of responders in e**xperimental group** (pelabresib + ruxolitinib) who achieve ≥50% reduction in total symptom score (TSS response) at Week 24

$P2_C$: proportion of responders in **control group** (placebo + ruxolitinib) who achieve ≥50% reduction in total symptom score (TSS response) at Week 24

For the main analysis, Cochran-Mantel Haenszel test using the stratification factors as per randomization on ITT will be used to compare SVR35 response rates between two arms.

Other Endpoints

For all other endpoints, inference testing may be performed, and associated p-values will be presented for descriptive purposes only.

6.1.5.2 Multiplicity Adjustment

The hypotheses testing will be adjusted for multiple comparisons using **the fixed sequence testing** procedure to control the family-wise Type 1 error rate. The pre-specified hypothesis tests for the primary endpoint SVR35 at Week 24, the key secondary endpoint absolute change in TSS at Week 24 compared to baseline, and the key secondary endpoint TSS50 at Week 24 will be performed in a sequential manner as shown in Figure 2. First, the hypothesis test for the primary endpoint will be conducted at two-sided $\alpha = 5\%$ level. If the null hypothesis for the primary endpoint is rejected in favor of the alternative hypothesis, the hypothesis test for the key secondary endpoints will subsequently be conducted at two-sided $\alpha = 5\%$ level. If there is no evidence to reject the null hypothesis for the primary endpoint, the subsequent key secondary endpoint tests will be performed, and p-values will be presented for descriptive purposes. Second, If the null hypothesis for the key secondary endpoint absolute change of TSS at Week 24 is rejected in favor of the alternative hypothesis, the hypothesis test for the TSS50 at Week 24 will subsequently be conducted at two-sided $\alpha = 5\%$ level. If there is no evidence to reject the null hypothesis for the key secondary endpoint absolute change in TSS at Week 24, the subsequent key secondary endpoint TSS50 at Week 24 test will be performed, and p-values will be presented for descriptive purposes. This fixed-sequence testing procedure ensures that the overall type I error rate for testing all endpoints is maintained at the 5% significance level.

6.2 Patient Disposition

A listing of all Screen Failures (i.e., patients who were screened but not randomized) will be presented along with reasons for screen failure and details of any AEs. Patients who were dosed but not enrolled will be marked in the listing.

The number of patients included in each analysis set will be summarized by treatment group and for the overall population. Percentages will be provided using the ITT Set as the denominator.

The number and percentage of patients in the following categories will be presented by treatment group and overall for all patients screened.

- Screen failures
- Patients randomized, but not treated

- Inclusion and/or exclusion criteria not met but randomized
- Patient status (i.e., ongoing, crossed over, discontinued) at the data cutoff
- Patients in each Analysis Sets
- Patients who received both study drug component
- Patients who received only pelabresib/placebo
- Patients who received only ruxolitinib
- Patients who discontinued study with reasons
- Patients discontinued double-blind treatment with reasons
- Patients discontinued crossover treatment with reasons
- Patients who discontinued all study drugs and ongoing in study at the data cut-off date
- Patients who had completed 24 weeks treatment
- Overall study duration using KM estimate

6.3 Protocol Deviations

Patients will be examined on a case-by-case basis to determine if conditions set forth in the study protocol have been violated. Protocol deviations related to COVID-19 pandemic will be captured. The CRO/sponsor will classify major/significant and minor/non-significant protocol deviations per the agreed Protocol Deviation Specification and PD review process. The study team will review the protocol deviations and their classification throughout the study.

For any criteria for protocol deviations that can be completely implemented by a computer program, the detailed algorithm will be agreed upon. Details of such algorithms will be included in the derived dataset specifications and finalized before treatment unblinding. Non-programable protocol deviations identified by medical monitors, medical data reviewer, CRA, and sponsor will be incorporated into the datasets.

Confirmed major and minor protocol deviations will be documented in the Protocol Deviation tracker for the study.

Major protocol deviations will be summarized and listed for the ITT Set for each treatment group. A summary and listing of PDs causing patients to be excluded from the PP set will be provided.

All protocol deviations in relation to COVID-19 will be identified based on reviews of the data prior to database lock. All major protocol deviations related to COVID-19 will be listed.

The number and percentage of all randomized patients, with at least one of important protocol deviation in relationship to COVID-19, will be summarized by treatment arm and overall according to the earlier described PD categories.

6.4 Demographic and Baseline Characteristics

6.4.1 Demographics

Descriptive summaries of demographic and baseline characteristics will be presented by treatment group and overall, for the ITT Set.

The following baseline demographics will be summarized:

- Age (at time of consent) in years

- Age group (at time of consent): 18 - 44 years, 45 - 64 years, ≥ 65 years
- Weight (kg), Height (cm) and Body Mass Index (kg/m^2) at baseline
- Race
- Ethnicity
- Sex
- Country
- Region

A listing will be created to show all the demographics and baseline characteristics for each patient in the ITT Set.

Additional Study Level Summary Statistics

A separate table will be generated to describe the maturity of data including the following:

- Duration of randomization (recruitment) period = (Date of last patient randomized - Date of first patient randomized + 1)/30.4375 (months)
- Date of the first patient randomized
- Date of the last patient randomized
- Cut-off date
- Duration of study follow up period

6.4.2 Disease Characteristics at Baseline

Baseline characteristics including MF disease history will be summarized using ITT Set. The following baseline characteristics will be treated as categorical variables:

- Myelofibrosis subtype: PMF (Primary Myelofibrosis), PV-MF (Post-Polycythemia Vera Myelofibrosis), and PET-MF (Post-Essential Thrombocythemia Myelofibrosis)
- ECOG performance status score: 0, 1, 2, ≥ 3
- DIPSS risk category: Intermediate-1 vs. Intermediate-2 vs. High
- Platelet count category: $> 200 \times 10^9$/L vs. $\leq 200 \times 10^9$/L
- Spleen volume category: ≥ 1800 cm^3 vs. < 1800 cm^3
- Baseline mutation status

The following baseline characteristics will be treated as continuous variable:

- Hemoglobin (g/dL)
- Platelet count
- Spleen volume (cm^3)
- Total Symptom Score using the MFSAF v4.0
- Time since MF diagnosis (months)

In addition, a table of discordance between stratification factors collected in IRT vs. baseline characteristics per eCRF will be created. Discrepancies between randomization stratification factors and baseline factors collected on eCRFs may be tabulated and listed.

- Randomization MF disease history (DIPSS) in IRT vs. MF Disease History (DIPSS) in CRF
- Spleen volume (cm^3) at randomization in IRT vs. baseline spleen volume per MRI/CT Imaging scan using local read
- Platelet count at randomization in IRT vs. Baseline platelet count in CRF

6.5 Medical History and Concomitant Medications

The medical history will be summarized by system organ class (SOC) and preferred term (PT) for each treatment group and overall, for the Safety Set.

6.5.1 Prior, Concomitant, and Post-Treatment Medications

World Health Organization (WHO) Drug Dictionary March 2021 will be used for coding prior and concomitant medications, classified by Anatomical Therapeutic Chemical (ATC) class and preferred drug name.

Prior medication is defined as any medication (therapies/procedures) with the start date and end date prior to the date of the first dose of study treatment.

Concomitant medication is defined as any medication with a start date prior to the date of the first dose of study treatment and continuing after the first dose of study treatment or with a start date and time after the study treatment initiation and before the end of the on-treatment period. The on-treatment period starts at the time of study treatment initiation through 7 days after the last dose of study treatment.

Any other permitted medication or vaccine (including over-the-counter or prescription medicines, vitamins, and/or herbal supplements) that the patient is receiving at the time of enrollment or receives during the study must be recorded along with:

- Reason for use
- Dates of administration including start and end dates
- Dosage information including dose and frequency

Post-treatment medication is defined as any medication with a start date after the end of the on-treatment period through the end of safety follow-up period.

Prior and concomitant medication usage will be summarized by the number and proportion of patients in each treatment group receiving each medication by ATC Level 4 and preferred term by treatment group and overall using the Safety Set. Medications can be counted both as prior and concomitant medication. Multiple medication usage by a patient in the same category will be counted only once.

All prior, concomitant, and post-treatment medication will be provided by patient in the listings.

6.6 Efficacy Analysis

All efficacy analyses will be conducted according to the treatment assigned.

All statistical tests will be 2-sided hypothesis tests performed at the 5% level of significance. All confidence intervals will be 2-sided 95% confidence intervals, unless stated otherwise.

6.6.1 Primary Endpoint Analysis (SVR35 at Week 24)

The primary endpoint of the study is splenic response defined as a ≥ 35% reduction from baseline in spleen volume as measured by MRI or CT at Week 24 and assessed by central radiology read.

6.6.1.1 Derivation of Endpoint

At Week 24 per visit window (Section 9.5), patients with a ≥ 35% spleen volume reduction (SVR) will be considered as responders, and patients without a ≥ 35% SVR, including missing assessments, will be considered as non-responders. Derivation as follows:

- SVR at Week n = Baseline spleen volume - Spleen volume at Week n
- Percentage of SVR = SVR/Baseline spleen volume * 100%

Patients will be classified as responders if the percentage of SVR ≥ 35% at Week 24.

If a patient dies, starts a new anti-MF treatment as per the Subsequent Anticancer Therapies CRF, or has disease progression as per the Disease Status Assessment CRF prior to Week 24 (Day 168 post randomization), then this patient will be considered as a "non-responder" irrespective of this MRI assessment.

Centrally assessed spleen volume as measured by MRI or CT will be used for the analysis.

Baseline definition is provided in Section 4.1.8. Details on the analysis visit windows are provided in Section 9.5.

6.6.1.2 Main Analytical Approach

The difference in proportion of responders between Experimental Group and Control Group will be obtained using Cochran-Mantel-Haenszel (CMH) weighted average across all strata, with baseline values of DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), platelet count (100-200 × 10^9/L vs. > 200 × 10^9/L), and spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) from IRT randomization as three stratification factors. The 95% confidence intervals (CI) of the weighted average of difference across strata will be provided using the normal approximation. If the minimum observed number of patients in a response category in a treatment group within a level of one of the stratification factors is less than 6, then the adjacent groups will be combined into 1 stratum level. The CMH weighted 95% CI is adjusted across the strata and this will form the basis of hypothesis testing. As supportive data an unadjusted 95% confidence interval based on combining the strata will be reported based on the normal approximation to the binomial distribution.

Homogeneity across strata will be tested using Breslow-Day test. If the test is significant at the alpha=0.05 level, stratum-specific differences in proportions will be reported and the adjusted difference and 95% CI in proportion of responders between treatment groups will be obtained and tested using minimum risk weight method of Mehrotra and Railkar (2000) instead of the CMH method, stratifying by baseline DIPSS, Platelet count and spleen volume.

Descriptive Reporting

Descriptive data will be presented for each treatment group and overall. The following will be provided:

- The number and percentage of patients achieving a ≥ 35% reduction from baseline in spleen volume at Week 24 will be summarized; Odds ratios (ORs), and 95% confidence limits of the binomial distribution will be provided.
- A stratified logistic regression model will be used to compare SVR35 rates between two treatment groups. Adjusted odds ratios (ORs) and corresponding 95% confidence intervals (CIs) and p-values will be estimated controlling for baseline scores.

- The percent change from baseline to Week 24 in spleen volume will be summarized and it will also be presented using a waterfall plot. The waterfall plot will present distribution bar chart and will be sorted in descending order.

Statistical Hypothesis Testing

Formal statistical hypothesis testing will be conducted as detailed in Section 6.1.5.1.

6.6.1.3 Sensitivity and Supplementary Analyses

The following sensitivity and supplementary analyses as specified in Table 2 will be performed for the primary efficacy endpoint.

Table 2: Sensitivity and Supplementary Analyses on SVR35

Type	Analysis
Sensitivity Analysis 1	Analysis to assess non-Responder due to missing data (ITT Set)
Sensitivity Analysis 2	Analysis using LOCF (ITT Set)
Sensitivity Analysis 3	Analysis using Multiple Imputation (ITT Set)
Sensitivity Analysis 4	Analysis using stratification level from randomization (ITT Set)
Supplementary Analysis 1	Analysis of SVR35 at Week 24 (PP Set)
Supplementary Analysis 2	Analysis of SVR35 at Week 24 using local imaging record (ITT Set)
Supplementary Analysis 3	Analysis of SVR35 at Week 24 (mITT Set)
Concordance Analysis	Concordance Rate between Central and Local imaging records on SVR35

6.6.1.3.1 Sensitivity Analyses

Some patients classified as SVR35 Week 24 non-responders due to missing data will be re-classified as responders or non-responders in some sensitivity analyses.

- **Sensitivity Analysis 1: Analysis of SVR35 to assess non-Responder due to missing data**

a. Rationale

For patients with missing spleen volume assessment during the Week 24 visit window, the primary analysis follows a derivation process for the responder status as specified in Section 6.6.1.1. To assess potential impact of the derivation and imputation process, an alternative imputation rule that does not necessarily account all missing values as non-responders will be performed.

b. Derivation of Endpoint

Same as primary analysis, except that patient classified as non-responders due to missing data in the primary analysis will be classified as responders in this sensitivity analysis if they meet the following:

- Patients discontinue treatment prior to Week 24 spleen assessment window, and have a SVR $\geq$ 35% at the time of treatment discontinuation

- Patients miss spleen volume assessment during the Week 24 window but complete an MRI or CT spleen size imaging with SVR ≥ 35% prior to Week 36.

Patients who discontinue the treatment phase early and take a new anti-cancer therapy prior to discontinuation will be classified as non-responders. Patients who discontinue the treatment phase early due to reasons other than initiation of new anti-cancer therapy will have response determined based on spleen volume assessments at the time of early discontinuation.

c. Analytical Approach

The criteria for the primary analysis outlined in Section 6.6.1.2 will be followed.

- **Sensitivity Analysis 2: Analysis of SVR35 using LOCF**

This sensitivity analysis will be conducted only if there are ≥10% of patients classified as non-responders due to missing data in the primary analysis.

a. Rationale

This sensitivity analysis is proposed to replace derivation rules in Section 6.6.1.1. by applying with last observation carry-forward (LOCF) for patients with missing Week 24 spleen values. The latest spleen volume assessment prior to Week 24 will replace the missing values at Week 24.

b. Derivation of Endpoint

The latest spleen volume assessment prior to Week 24 will replace the missing values at Week 24, except for death or start of a new anti-cancer therapy or Progressive disease prior to Day 168, when patients will be considered as non-responders.

c. Analytical Approach

The criteria for the primary analysis outlined in Section 6.6.1.2 will be followed.

- **Sensitivity Analysis 3: Analysis of SVR35 using Multiple Imputations**

This sensitivity analysis will be conducted only if there are ≥10% of patients classified as non-responders due to missing data in the primary analysis.

a. Rationale

This sensitivity analysis is proposed to replace derivation rules in Section 6.6.1.1. by first applying multiple imputations for missing spleen values and then applying the response algorithm at Week 24.

b. Derivation of Endpoint

Missing spleen value will be imputed using a parametric regression model with the assumption of multivariate normality and a monotone missing data pattern. If there are patients with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using the Markov Chain Monte Carlo method.

Before missing value handling, all spleen values after the discontinuation of the double-blind treatment will be imputed as missing. Visit windows will be used to designate one spleen value per patient per visit from Baseline to Week 24. Visit windows and rules for selecting or calculating the single spleen value are described in Section 9.5. Data will only be imputed for a visit if there is no available data at that visit and the study day of the imputed data will be the scheduled day of the visit.

The missing values of spleen volume will be imputed as follows, creating 50 imputed datasets.

First, values for patients with a non-monotone missingness pattern will be filled, by assuming a multivariate normality over the values of spleen volume at all time points. The result of this imputation will be 50 datasets for which the values of non-missing data are identical, the values of the non-monotone missing data are draws, and the remaining missingness pattern is monotone.

Next, a sequential regression approach will be used to fill in the monotone missing data. The DIPSS risk category, baseline platelet count as continuous variable, baseline spleen volume as continuous variable, the planned treatment group, and baseline TSS will be used as covariates in the imputation.

Week 12 missing spleen values will be filled in using regression of non-missing spleen volume values on the listed covariates. Week 24 missing spleen values will be filled in using regression on the listed covariates and the Week 12 spleen values.

Once missing data are filled in, the imputed data sets will be converted to report Week 24 responder status. After applying the response algorithm, the dataset will be analyzed using the CMH test for the adjusted difference in proportion of response controlling for three stratification factors, as in the primary analysis.

c. Analytical Approach

The criteria for the primary analysis outlined in Section 6.6.1.2 will be followed.

- **Sensitivity Analysis 4: Use of Stratification Level from Randomization**

If there is a high rate of discrepancy (>10%) between the strata classifications from clinical database which were integrated from the IRT randomization data and the baseline strata, below sensitivity analysis will be performed. The strata variables used for this analysis will be derived efficacy baselines based on the data collected in the clinical database.

a. Rationale

The baseline values of DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), platelet count (100-200 $\times 10^9$/L vs. > 200 $\times 10^9$/L), and spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) are used to determine the stratum in the primary analysis. However, the randomization stratification baselines were based on the screening from local laboratory. An analysis will be performed to reflect the actual stratification.

b. Derivation of Endpoint:

Same as primary analysis

c. Analytical Approach

The analysis of the primary efficacy endpoint will be repeated using the derived efficacy baselines – last assessments prior to randomization. This is to evaluate the impact of treatment effect on primary analysis due to the difference in stratum assignment used in randomization and the baseline laboratory results.

6.6.1.3.2 Supplementary Analyses

The following supplementary analyses will be conducted for primary endpoints to evaluate the robustness of the results from the primary analysis methods:

1. **Analysis of SVR35 at Week 24 using PP Set**

   The same analytical approach as Section 6.6.1.2, imputing missingness as non-responders; Analyses will be based on PP Set.

2. **Analysis of SVR35 at Week 24 using local imaging record**

The same analytical approach as Section 6.6.1.2, imputing missingness as non-responders; Analyses will be based on local imaging record.

3. Analysis of SVR35 at Week 24 using mITT Set

The same analytical approach as Section 6.6.1.2, imputing missingness as non-responders; Analyses will be based on mITT Set.

6.6.1.3.3 Concordance Analyses

Concordance Rate between Central and Local Imaging Records on SVR35 at Week 24

The concordance between the number of SVR35 responders and non-responders at Week 24 in the central versus the local records will be calculated for both treatment groups. This analysis will be conducted only on patients with a SVR35 from both central and local imaging data at Week 24 - hence the concordance rate will be computed based on non-missing data. The concordance table, however, will show missing data as well (see below).

The central and local imaging SVR data records will be 1-to-1 matched. Afterwards, the concordance rate will be computed as indicated below.

Table 3: Computation of the Concordance Rate between Central and Local Imaging SVR Data Records

		Local			
Central		Responder	Non-Responder	Total Non-Missing	Missing
	Responder	a	b	a+b	g
	Non-Responder	c	d	c+d	h
	Total Non-Missing	a+c	b+d	a+b+c+d	g+h
	Missing	e	f	e+f	
		Concordance rate		$100 * \frac{(a+d)}{(a+b+c+d)}$	

6.6.1.3.4 Subgroup Analyses

The following subgroup analyses will be conducted per treatment group and overall, for the ITT Set:

The percent reduction in spleen volume from Baseline to Week 24 will be graphically reported for the subgroup categories as specified in Table 4 below using a forest plot. For each subgroup category, a forest plot will display the mean percent reduction in spleen volume from Baseline to Week 24 as well as the corresponding 95% confidence interval.

The splenic response rate at Week 24 will be graphically displayed for the subgroup categories as specified in Table 4 using a forest plot. For each subgroup category, the forest plot will show the splenic response rate at Week 24 together with the corresponding 95% confidence interval.

Table 4: Subgroup Categories

Subgroup
Gender (Female, Male)
Age Group (≥ 65 years, < 65 years)
Race Group

Subgroup
Ethnicity
Region
DIPSS (intermediate-1risk, intermediate-2, and High risk)
MF Subtype (PMF, PPV-MF or PET-MF)
Baseline Spleen Volume (≥ 1800 cm^3, < 1800 cm^3)
Baseline Platelet Count (> 200 × 10^9/L, ≤ 200 × 10^9/L)
Baseline Hemoglobin (> 10 g/dL, ≤ 10 g/dL)
Time since diagnosis (months) (< Median, ≥ Median)

6.6.2 Key Secondary Endpoint Analysis (Absolute Change from Baseline TSS at Week 24)

The first key secondary endpoint of the study is absolute change from baseline TSS at Week 24.

6.6.2.1 Derivation of Endpoint

At Week 24, the absolute change from baseline TSS will be calculated. Derivation as follows:

- Daily TSS = Sum of the 7 items on the MFSAF v4.0 form for a given day
- Baseline TSS = Averaged daily TSS from the 7-day period prior to randomization (Day -7 to Day -1, where Day -1 is the day on or prior to randomization depending on the actual randomization time. Denominator will be number of days with non-missing TSS score)
- Week 24 TSS = Averaged daily TSS at Week 24
- Change from baseline TSS at Week 24 = Week 24 TSS – Baseline TSS

If there are less than 4 daily MFSAF available for Week 24 then Week 24 TSS will be set to missing. Missing TSS at Week 24 will be imputed using Multiple Imputation (MI).

The missing weekly TSS will be imputed as follows, creating 50 imputed datasets.

First, values for patients with a non-monotone missingness pattern will be filled, by assuming a multivariate normality over the weekly TSS at all weeks. The result of this imputation will be 50 datasets for which the values of non-missing data are identical, the values of the non-monotone missing data are draws, and the remaining missingness pattern is monotone.

Next, a sequential regression approach will be used to fill in the monotone missing data. The DIPSS risk category, baseline platelet count as continuous variable, baseline spleen volume as continuous variable, the planned treatment group, and baseline TSS will be used as covariates in the imputation.

Week 24 missing TSS values will be filled in using regression on the listed covariates and the Week 1 through Week 23 TSS values. The weekly TSS will be based on section 9.6

6.6.2.2 Main Analytical Approach

Descriptive summary statistics including the number of patients (n), mean, standard deviation, median, Q1, Q3, minimum, and maximum of TSS values at Baseline and Week 24, and the absolute change from baseline at Week 24 will be present.

An ANCOVA (Analysis of Covariance) model will be used to analyze the absolute change from baseline in

TSS at Week 24. The dependent variable is change from baseline in TSS at Week 24, with treatment group as independent variable, baseline DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), baseline platelet count (100-200 × 10^9/L vs. > 200 × 10^9/L) and baseline spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) and the baseline TSS as covariates.

Implementation of MI + ANCOVA in SAS are described below.

Step 1: Impute missing TSS using SAS PROC MI:

- PROC MI with option NIMPUTE=0 will used to exam the missing patterns.
- For patients with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using the Markov Chain Monte Carlo (MCMC) method.
- Impute remaining data with regression on above partially imputed output from MCMC with monotone missing data pattern.

Step 2: Calculate change from baseline TSS at Week 24 and perform ANCOVA on each imputed dataset using SAS PROC MIXED to obtain:

- P-values for the overall treatment effect at Week 24
- LSM (Least Squares Means) estimates for the change from baseline in TSS at Week 24 for each treatment group with standard errors (SEs)
- LSM (Least Squares Means) estimates for difference in the change from baseline in TSS between experimental group and control group

Step 3: Combine ANCOVA results from the multiple imputed datasets using SAS PROC MIANALYZE

The LSM line plot for the combined inference will be presented for Week 24, with vertical bars representing ± one SE from the corresponding point estimate of the LS mean.

Box plots of absolute TSS change from baseline in TSS at Week 24 will also be provided.

6.6.2.3 Sensitivity and Supplementary Analyses

The following sensitivity and supplementary analyses as specified in Table 5 will be performed. Sensitivity Analyses 1 and 2 will be performed if data warranted.

Table 5: Sensitivity and Supplementary Analyses on Absolute Change from Baseline TSS

Type	Analysis
Sensitivity Analysis 1	Tipping Point Analysis of Absolute Change from Baseline in TSS at Week 24 (ITT Set)
Sensitivity Analysis 2	Analysis of Absolute Change from Baseline in TSS at Week 24 using Jump to Reference Imputation (ITT Set)
Supplementary Analysis 1	Analysis of Absolute Change from Baseline TSS using Mixed model Repeated Measures (MMRM) (ITT Set)

- **Sensitivity Analysis 1: Tipping Point Analysis of Absolute Change from Baseline TSS**

The sensitivity analysis 1 will be conducted by searching for a tipping point that reverses the study conclusion.

a. Rationale

The conclusion in Main Analytical Approach is based on the Missing at Random (MAR) assumption. However, this assumption might not be accurate. To further assess the different missing mechanism such as Missing Not at Random (MNAR), tipping point analysis will be used to identify the scenario where the treatment effect in subjects with missing data overturns the significant treatment effect obtained from analysis based on MAR.

b. Derivation of Endpoint

The calculation for absolute change from baseline in TSS at Week 24 is the same as in Section 6.6.2.1. A specified sequence of shift parameters will be applied with multiple imputation to adjust the imputed values. Tipping point will be identified where treatment effect in patients with missing Week 24 TSS data overturns the significant treatment effect obtained (p-value ≤ 0.05) in ANCOVA from main analytical approach.

c. Analytical Approach

The criteria for main analytical approach of the key secondary endpoint outlined in Section 6.6.2.2 will be followed. If the sensitivity analysis reveals that the "tipping point" is an unreasonable shift, then the robustness of the study result under the MAR assumption are supported.

- **Sensitivity Analysis 2: Jump-to-Reference Analysis**

For jump-to-reference imputation method, missing values in the pelabresib + ruxolitinib group following an intercurrent event are imputed using the distribution of the placebo + ruxolitinib group (the reference). The SAS macro, "RMConjPLUS", developed by James Roger (Carpenter et al. 2013; White et al. 2020) will be used.

a. Rationale

The standard MI assumed that patients from the experimental group continue to reap benefits from the experimental treatment after discontinuation. However, this is a study with a standard-of-care control treatment (ruxolitinib) and where most patients discontinued from the experimental group may continue their standard-of-care treatment. Therefore, it might be reasonable to assume that after withdrawal from the double-blind treatment, patients from the experimental group (no longer receiving pelabresib) will exhibit the same future evolution of the disease as patients on the control group. Patients that discontinue from the control group are assumed to evolve in the same way as control patients that remain in the study.

b. Derivation of Endpoint

The calculation for absolute change from baseline in TSS at Week 24 is the same as in Section 6.6.2.1. The absolute change from baseline in TSS at Week 24 from reference group will be used as dependent variable. DIPSS risk category, baseline platelet count (continuous), baseline spleen volume (continuous), weekly visit, DIPSS and weekly visit interaction, baseline platelet count (continuous) and weekly visit interaction, baseline spleen volume (continuous) and weekly visit interaction, baseline TSS and weekly visit interaction will be used as covariates in the imputation model.

c. Analytical Approach

LSM (Least Squares Means) estimates for the change from baseline in TSS at Week 24 for each treatment group with standard errors (SEs) using ANCOVA model will be calculated. In the ANCOVA model, the dependent variable is change from baseline in TSS at Week 24, with treatment group as independent variable, baseline DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), baseline platelet count (100-200 $\times 10^9$/L vs. $> 200 \times 10^9$/L) and baseline spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) and the baseline TSS as covariates. LSM (Least Squares Means) estimates for difference in the change from baseline at Week 24 in TSS between pelabresib + ruxolitinib group and placebo + ruxolitinib group with associated p-value will also be calculated.

- **Supplementary Analysis 1: Analysis of Absolute Change from Baseline TSS using Mixed model Repeated Measures (MMRM)**

A MMRM model will be used to analyze the absolute change from baseline for the 7-day average TSS every 4 weeks using treatment group, analysis timepoint (Week 4, 8, 12, 16, 20, 24), treatment group-by-every 4-week interaction, baseline DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), baseline platelet count (100-200 × 10^9/L vs. > 200 × 10^9/L) and baseline spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) as factors and baseline 7-day average TSS as a covariate. An unstructured covariance matrix will be used to model the within-patient variance-covariance errors. If the unstructured covariance structure matrix results in a lack of convergence, the heterogenous first-order autoregressive covariance structure will be used as the first alternative, followed by the compound symmetry covariance structure. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Any missing data are assumed to be missing at random.

The LSM for each treatment group with SE, LSM difference against control group (placebo + ruxolitinib) with SE, 95% CI for the differences and the associated p-values will be presented at each analysis timepoint. The overall treatment effect 24 weeks will also be estimated.

6.6.2.4 Subgroup Analyses

ANCOVA of the absolute change from baseline TSS at Week 24 will be performed using subgroups defined in Table 4. For each subgroup category, a forest plot will display the mean reduction in TSS from baseline to Week 24 as well as the corresponding 95% confidence interval.

6.6.3 Key Secondary Endpoint Analysis (TSS50 at Week 24)

The second key secondary endpoint of the study is TSS response, defined as a ≥ 50% decrease from baseline in TSS as measured by the MFSAF v4.0, at Week 24.

6.6.3.1 Derivation of Endpoint

At Week 24, Patients with a ≥ 50% decrease from baseline in TSS as measured by the MFSAF v4.0 will be considered as responders, and patients without a 50% decrease from baseline in TSS as measured by the MFSAF v4.0, including missing assessments, will be considered as non-responders.

At Week 24 per visit window (Section 9.7), the percent change from baseline TSS will be calculated. Derivation as follows:

- Daily TSS = Sum of the 7 items on the MFSAF v4.0 form for a given day
- Baseline TSS = Averaged daily TSS from the 7-day period prior to randomization (Day -7 to Day -1, where Day -1 is the day on or prior to randomization depending on the actual randomization time. Denominator will be number of days with non-missing TSS score)
- Week 24 TSS = Averaged daily TSS at Week 24 per visit window (Section 9.7)
- Change from baseline TSS at Week 24 = Week 24 TSS – Baseline TSS
- Percent change from baseline TSS at Week 24 = Change from baseline TSS at Week 24/Baseline TSS *100%

Daily TSS is a 24-hour recall (i.e., daily diary) format of the MFSAF v4.0, which equals the sum of the 7 individual item responses on the 0-10 scale, with a possible total daily score that may range from 0 to 70. For each individual question, the response scale is 0 (Absent) to 10 (Worst Imaginable). All 7 items must be completed for a daily TSS to be computed. If missing response for any of 7 questions on the daily diary, the TSS will be considered as missing for that day.

Baseline TSS is calculated as the average of non-missing daily total symptom scores over the 7- day period prior to day of randomization. The baseline TSS will be considered as missing if there are no TSS for 7- day period prior to day of randomization.

Patients will be classified as responders if percentage of change from baseline TSS is ≤ -50% at Week 24 per visit window in Section 9.7.

Patients will be classified as non-responders if any of the following occurred:

1. Percentage of change from baseline TSS > -50% at Week 24
2. Missing Week 24 TSS after adjusted per analysis visit window
3. Death, start of a new anti-cancer therapy or progressive disease prior to Week 24

6.6.3.2 Main Analytical Approach

The difference in proportion of responders between Experimental Group and Control Group will be obtained using Cochran-Mantel-Haenszel (CMH) weighted average across all strata, with baseline DIPSS (Intermediate-1 vs. Intermediate-2 vs. High), platelet count (100-200 × 10^9/L vs. > 200 × 10^9/L), and spleen volume (< 1800 cm^3 vs. ≥ 1800 cm^3) from IRT randomization as three stratification factors. The 95% confidence intervals (CI) of the weighted average of difference across strata will be provided using the normal approximation. If the minimum observed number of patients in a response category in a treatment group within a level of one of the stratification factors is less than 6, then the adjacent groups will be collapsed into 1 stratum level. The CMH weighted 95% CI is adjusted across the strata and this will form the basis of hypothesis testing. As supportive data an unadjusted 95% confidence interval based on combining the strata will be reported based on the normal approximation to the binomial distribution.

Homogeneity across strata will be tested using Breslow-Day test. If the test is significant at the alpha = 0.05 level, stratum-specific differences in proportions will be reported and the adjusted difference and 95% CI in proportion of responders between treatment groups will be obtained and tested using minimum risk weight method of Mehrotra and Railkar (2000) instead of the CMH method, stratifying by baseline DIPSS, Platelet count and spleen volume.

Main analysis will be conducted based on the ITT Set.

Descriptive Reporting

Descriptive data will be presented for each treatment group and overall. The following will be provided:

- The number and percentage of patients achieving a ≥ 50% decrease from baseline TSS at Week 24 will be summarized; Odds ratios (ORs), 95% confidence limits of the binomial distribution, and p-values will be provided.
- A stratified logistic regression model will be used to compare TSS50 rates between two treatment groups. Adjusted odds ratios (ORs) and corresponding 95% confidence intervals (CIs) and p-values will be estimated controlling for baseline scores.
- The percent change from baseline to Week 24 in TSS will be summarized and it will also be presented using a waterfall plot. The waterfall plot will present distribution bar chart and will be sorted in descending order.

Statistical Hypothesis Testing

Formal statistical hypothesis testing will be conducted as detailed in Section 6.1.5.1.

6.6.3.3 Sensitivity and Supplementary Analyses

The following sensitivity and supplementary analyses as specified in Table 6 will be performed.

Table 6: Sensitivity and Supplementary Analyses on TSS50 at Week 24

Type	Analysis
Sensitivity Analysis 1	Analysis to assess non-Responder due to missing data (ITT Set)
Sensitivity Analysis 2	Analysis using LOCF (ITT Set)
Sensitivity Analysis 3	Analysis using Multiple Imputation (ITT Set)
Sensitivity Analysis 4	Analysis using stratification level from randomization (ITT Set)
Supplementary Analysis 1	Analysis of TSS50 at Week 24 (PP Set)
Supplementary Analysis 2	Analysis of TSS50 at Week 24 (mITT Set)
Concordance Analysis	Concordance Rate between SVR35 and TSS50 Responders/non-Responders

Sensitivity Analyses of TSS50 at Week 24

Some patients classified as TSS50 Week 24 non-responders due to missing data will be re-classified as responders or non-responders in some sensitivity analyses.

- **Sensitivity Analysis 1: Analysis of TSS50 to Assess non-Responder due to missing data**

a. Rationale

For patients with incomplete or missing MFSAF v4.0 responses, the daily TSS will be set as missing. For a given week, if a patient has more than 3 days of missing TSS, the weekly-averaged TSS will be set as missing. Patients will be counted as non-responders if missing the weekly-averaged TSS at Week 24 per visit window in Section 9.7. To assess potential impact of the derivation and imputation process, an alternative imputation rule that does not necessarily account all missing values as non-responders will be performed.

b. Derivation of Endpoint

Patients with incomplete MFSAF v4.0 will be handled as following:

- For incomplete MFSAF forms with no more than 2 items missing, the missing fields will be imputed by the maximum of the 5 or 6 non-missing scores. Daily TSS = sum of the 7 scores after imputation. If there are 3 or more items missing, the daily TSS will be missing.

The TSS50 response derivation is the same as the main analysis, except below:

- Remove the 4 out of 7-day daily MFSAF requirement. The weekly averaged TSS will be based on any non-missing scores completed.

c. Analytical Approach

The criteria for main analytical approach of the key secondary endpoint outlined in Section 6.6.3.2 will be followed.

- **Sensitivity Analysis 2: Analysis of TSS50 using LOCF**

This sensitivity analysis will be conducted only if there are ≥10% of patients classified as non-responders due to missing data in the primary analysis.

a. Rationale

This sensitivity analysis is proposed to replace derivation rules in Section 6.6.3.1 by applying LOCF for patients with missing Week 24 TSS. The latest weekly TSS prior to Week 24 (after visit window adjustment) will replace the missing Week 24 TSS.

b. Derivation of Endpoint

The latest weekly TSS prior to Week 24 will replace the missing values at Week 24, except for death, progressive disease, or transition to a new anti-cancer therapy, which missing will be set as non-responders.

c. Analytical Approach

The criteria for main analytical approach of the key secondary endpoint outlined in Section 6.6.3.2 will be followed.

- **Sensitivity Analysis 3: Analysis of TSS50 using *Multiple Imputations***

This sensitivity analysis will be conducted only if there are ≥10% of patients classified as non-responder due to missing data in the primary analysis.

a. Rationale

This sensitivity analysis is proposed to replace derivation rules in Section 6.6.3.1 by first applying with multiple imputations for missing TSS and then applying the response algorithm at Week 24.

b. Derivation of Endpoint

Missing weekly TSS will be imputed using a parametric regression model with the assumption of multivariate normality and a monotone missing data pattern. If there are patients with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using the Markov Chain Monte Carlo method.

Before missing value handling, all TSS values after the discontinuation of the double-blind treatment will be imputed as missing. Data will be imputed only for weeks in which TSS are missing.

The missing weekly TSS will be imputed as follows, creating 50 imputed datasets.

First, values for patients with a non-monotone missingness pattern will be filled, by assuming a multivariate normality over the weekly TSS at all weeks. The result of this imputation will be 50 datasets for which the values of non-missing data are identical, the values of the non-monotone missing data are draws, and the remaining missingness pattern is monotone.

Next, a sequential regression approach will be used to fill in the monotone missing data. The DIPSS risk category, baseline platelet count as continuous variable, baseline spleen volume as continuous variable, the planned treatment group, and baseline TSS will be used as covariates in the imputation.

Week 24 missing TSS values will be filled in using regression on the listed covariates and the Week 1 through Week 23 TSS values.

Once missing data are filled in, the imputed data sets will be converted to report Week 24 responder status. After applying the response algorithm, the dataset will be analyzed using the CMH test for the adjusted difference in proportion of response controlling for three stratification factors, as in the primary analysis.

c. Analytical Approach

The criteria for main analytical approach of the key secondary endpoint outlined in Section 6.6.3.2 will be followed.

- **Sensitivity Analysis 4: Use of Stratification Level from Randomization**

If there is a high rate of discrepancy (>10%) between the strata classifications from clinical database which were integrated from the IRT randomization data and the baseline strata, sensitivity analysis 4 will be performed. Same principal as "Sensitivity Analysis 4" in Section 6.6.1.3.1, the derived efficacy baselines (last assessment prior to randomization) will be used.

Supplementary Analyses

Supplementary analyses as listed in Table 6 will be conducted for TSS50 on the PP set and ITT set to evaluate the robustness of the results from the primary analysis methods. The same derivation of endpoint and analytical approach will be applied as described in Section 6.6.3.1 and Section 6.6.3.2.

Compliance Rate Summary

TSS compliance rate at each assessment is the number of patients with non-missing scores divided by the number of patients eligible for assessment at that analysis timepoint (eligible = ongoing until the end of the week).

A table and a figure will be presented for the weekly compliance rate by week and by treatment group.

Concordance Analyses

Concordance Rate between SVR35 and TSS50 Responders/non-Responders at Week 24

This concordance analysis is aimed at analysing the concordance between the number of responders and non-responders at Week 24 in SVR35 (based on central read) versus the MFSAF TSS50. The concordance rate will be computed only on patients with both SVR35 and TSS50 at Week 24 for both treatment groups. However, the concordance table will show missing data as well.

The data on SVR35 and TSS50 will be 1-to-1 matched. Afterwards, the concordance rate will be computed as indicated below.

Table 7: Concordance Rate between SVR35 and TSS50 Responders/non-Responders at Week 24

		SVR35 at Week 24			
MFSAF TSS50 at Week 24		Responder	non-Responder	Total non-Missing	Missing
	Responder	a	b	a+b	g
	non-Responder	c	d	c+d	h
	Total non-Missing	a+c	b+d	a+b+c+d	g+h
	Missing	e	f	e+f	
		Concordance rate		$100 * \frac{(a+d)}{(a+b+c+d)}$	

6.6.3.4 Subgroup Analyses

The following subgroup analyses will be conducted per treatment group and overall, for the ITT Set:

TSS50 response rate at Week 24 will be graphically displayed for the subgroup categories as specified in Table 4 using a forest plot. For each subgroup category, the forest plot will show the TSS50 response rate at Week 24 together with the corresponding 95% confidence interval.

6.6.4 Other Secondary Endpoints Analysis

The analysis of other secondary efficacy endpoints will be conducted using the ITT Set for the double-blind treatment period unless otherwise specified. Other secondary efficacy endpoints will be summarized by treatment group and overall.

6.6.4.1 Change from Baseline in TSS

TSS values at each analysis timepoint (every 12 weeks) will be calculated with window adjusted per Section 9.7. Missing data at each analysis timepoint will not be further imputed. Descriptive statistics will be generated for Baseline, Week 12, Week 24, Week 36, Week 48, and Week 60.

Descriptive summary statistics including the number of patients (n), mean, standard deviation, median, Q1, Q3, minimum, and maximum of TSS values at each analysis timepoint, and the absolute/percent change from baseline at each analysis timepoint will be presented.

Below analyses on percent Change from Baseline in TSS at Week 24 will be performed:

- The difference in CDF of percent change from baseline in TSS at Week 24 between the two treatment groups will be tested. P-value for the 2 sample Kolmogorov-Smirnov test will be reported. The difference in median percent change in TSS along with its 95% CI between the two treatment groups will be estimated using Hodges-Lehmann method and the p-value will be calculated using Wilcoxon rank sum test.
- The CDFs for each treatment group will be plotted with median.
- Box plots of the percent change from baseline in TSS at Week 24 will be provided.
- A waterfall plot of the percentage change in TSS at Week 24 will be presented by treatment. A reference line at -50% which indicates TSS50 response will be superimposed on the plot for reference.

Descriptive statistics and box plots of change from baseline of the 7 individual scores at each analysis timepoint will also be provided.

6.6.4.2 Change from Baseline in Bone Marrow Fibrosis at Week 24

Improvement in bone marrow fibrosis by at least 1 grade as assessed by central read compared to baseline will be analyzed by treatment group and overall. The improvement in bone marrow fibrosis grade is defined as a decrease by at least 1 grade in bone marrow fibrosis grade when compared to baseline, where a grade of MF-3 is the most severe, and MF-0 is the least severe.

- MF Grade (MF-0, MF-1, MF-2, MF-3)

The number and percentages of patients who meet the following categories will be summarized.

- Improved by at least 1 grade
- No Change
- Worsened

The presence of improvement by at least 1 grade (yes or no response) at Week 24 will be compared between treatment groups using a CMH test controlling for the randomization stratification factors. Besides, a stratified logistic regression model will be used to compare response rates between two treatment groups at Week 24

controlling for baseline scores. The ORs, nominal p-values, and corresponding 95% CIs will be derived.

A shift table will be produced presenting the number and percentage of patients in each bivariate category (i.e., baseline and Week 24) with regards to the result of the bone marrow fibrosis grade (i.e., grades 0-to-3). Percentages for the individual cells in the shift-table will be calculated based on baseline totals.

Bone marrow fibrosis grades from baseline to the best post-baseline result: the same shift table as above will be produced presenting number and percentage of patients in the bivariate categories baseline vs. best post-baseline result.

The distribution of bone marrow fibrosis change from baseline (i.e., number and percentage of patients in the above 3 categories) will be summarized for all available data points per visit window mapping in Section 9.8.

6.6.4.3 SVR35 at Week 12, Week 36, Week 48 and Week 60

Splenic response at Week n, defined as a ≥ 35% reduction from baseline in spleen volume as measured by MRI or CT and assessed by central radiology read, at Week n.

The criteria for the primary analysis outlined in Section 6.6.1.2 will be followed. By treatment group analysis will be based on the double-blind treatment period. All patients entering the crossover period prior to or on Week n will be considered as non-responders at Week n. Missing spleen values at Week n will be considered as non-responders except for patients who are ongoing and have not entered Week n yet. The endpoint derivation is the same as SVR35 at Week 24. Patients who are ongoing and have not entered Week 36, 48, 60 will be categorized as non-evaluable.

A separate table will be generated to present the number and percentage of patients who achieves SVR35 at Week 36, Week 48 and Week 60 for the crossover treatment period. Only patients entered the crossover cycle will be included in this summary.

Additional table and waterfall plots of the number and percentage of patients who achieved SVR35 at any time on the study (prior to death, progressive disease, or start of new anti-cancer treatment) will be provided.

6.6.4.4 TSS50 at Week 12, Week 36, Week 48 and Week 60

TSS response at Week n is defined as a ≥ 50% decrease from baseline in TSS as measured by the MFSAF v4.0, at Week n. The criteria for main analytical approach of the secondary endpoint TSS50 outlined in Section 6.6.3 will be followed. All patients entered the crossover period prior to or on Week n will be considered as non-responders for the double-blind treatment period. Missing TSS at Week n will be considered as non-responders, except for patients who are ongoing and have not entered Week 36, 48 and 60 which be categorized as non-evaluable. The endpoint derivation is the same as TSS50 at Week 24.

The compliance rate TSS50 at Week n will also be presented. The compliance rate defines as the total number of patients who achieve 4 out 7 TSS at Week n using analysis window defined in Section 9.6 divided by who are ongoing and have entered Week n.

A separate table will be generated to present the number and percentage of patients who achieves TSS50 at Week n in the crossover treatment period. Only patients entered the crossover cycle prior will be included in this summary.

Additional table and waterfall plots of the number and percentage of patients who achieved TSS50 at any time on the study (prior to death, progressive disease, or start of new anti-cancer treatment) will be provided.

6.6.4.5 Absolute Change from Baseline in TSS at Week 36, 48 and 60

The endpoint derivation is similar to the key secondary analysis outlined in Section 6.6.2.1, but for different analysis timepoint. Missing TSS values at Week n will be imputed using the same multiple imputation method as outlined in Section 6.6.2.2. Patients who are ongoing and have not entered Week 36, 48, 60 will be

excluded from the ANCOVA model. By treatment group analysis will be based on the double-blind treatment period.

A separate table will be generated to present the summary statistics for absolute/percent change from baseline in TSS at Week 36, Week 48, and Week 60 for the crossover treatment period. Only patients entered the crossover cycle will be included in this summary.

6.6.4.6 Rate of RBC Transfusions

Rate of RBC transfusions, defined as the average number of units of RBC transfusion per month (4 weeks), over the first 24 weeks of treatment.

The rate of RBC transfusion over the 24-week period will be derived for the RBC transfusion evaluable population, which are patients who have been on the study for 24-week without start of new anti-MF treatment. Patients who are not included in this analysis and the reason for being excluded will presented.

The total number of units transfused per 4-week period, starting from C1D1, will be derived for each patient. The baseline is derived as the 12-week period prior to C1D1.

The number of units of RBC transfusion per patient in each 4-week period and the total number of transfusions over 24 weeks will be summarized by treatment group.

The percentage change of units of RBC transfusion per patient in each 4-week period comparing to baseline and the total number of transfusions over 24 weeks comparing to baseline will be summarized by treatment group.

The rate of transfusion between the two treatment groups will be compared using 2-sample t-tests with pooled variances, and the difference between the treatment groups along with the 95% CI be estimated in terms of LS means.

A plot of average monthly rates of transfusion with 95% CI for the two treatment groups will be provided.

6.6.4.7 Conversion from RBC Transfusion Dependence

Transfusion dependence evaluable population are defined as:

- ≥6 units of RBC transfusion during the 12-week baseline period prior to dosing AND
- Patients with ≥6 units of RBC transfusion in the first 12 weeks OR
- Patients on study for at least 12 weeks

Patients discontinuing the study prior to 12 weeks receiving <6 units will be excluded from the evaluable set. Reasons for exclusion will be presented.

Conversion from RBC transfusion dependence (≥6 units of RBC transfusion during the 12-week baseline period prior to dosing) to independence (no RBC transfusions during a 12-week period of the double-blind treatment period).

The conversion rate from TD to TI is defined as the proportion of patients who converted from transfusion dependence (TD) to transfusion independence (TI), where:

TD is defined as: ≥6 units of RBC transfusion during the 12-week baseline period prior to dosing

TI is defined as: absence of RBC transfusions in any continuous 12-week period of the double-blind treatment period

Patients still in the double-blind treatment period and without a post-dose RBC transfusion in the latest 12 weeks will be considered TD to TI conversion responders. Patients who discontinued from the double-blind

treatment prior to Week 12 will be considered non-responders. Only patients who are TD at baseline will be included in the summary.

The number and percentage of patients who converted from TD to TI in the double-blind treatment period will be summarized; Odd's ration, 95% confidence limits of the binomial distribution will be provided.

A separate table will be created for patients converted from TD to TI in the crossover treatment period. TI in crossover is defined as no RBC transfusion in any continuous 12-week period of the crossover treatment period. Only patients entered the crossover cycle will be included in this summary.

6.6.4.8 PGIC at Week 24

The endpoint is category change of PGIC at Week 24. The PGIC is a single question to assess the patient's impression of change in their MF symptoms since the start of study treatment. The patient will answer the following question: "Since the start of the treatment you've received in this study, your myelofibrosis symptoms are (1) Very much improved, (2) Much improved, (3) Minimally improved, (4) No change, (5) Minimally worse, (6) Much worse, (7) Very much worse."

Patients are requested to complete the PGIC assessment electronically on a weekly basis after the start of study treatment irrespective of any delay in cycle due to drug hold. Patients should complete the PGIC on the same day each week until 12 weeks after the end of treatment.

The number and percentages of patients in each category per analysis timepoint in Table 16 will be tabulated. A bar plot of Week 24 symptom distribution will be provided.

The number of patients with at least 1 level improvement in PGIC, and the number of patients with at least 2 level improvement in PGIC at Week 24 will also be summarized.

6.6.4.9 Progression-Free Survival

6.6.4.9.1 Definition

PFS, defined as the time from randomization until documented progression, or until death from any cause for patients without documented progression, whichever comes first, will be analyzed.

Progressive disease is defined by meeting 1 of the following criteria:

- Progressive splenomegaly, defined as enlargement of spleen volume by MRI or CT scan of ≥ 25% compared to the baseline value, as confirmed by the central radiology review
- Leukemic transformation confirmed by a bone marrow blast count of ≥ 20% or
- A peripheral blood blast content of ≥ 20% associated with an absolute blast count of ≥ 1×10^9/L that lasts for at least 2 weeks

Table 8: Censoring Rules for PFS

Item	Situation	Date of Progression or Censoring	Outcome	Censoring reason reported
1	No post-baseline response assessment	Date of randomization	Censored	No post-baseline response assessment
2	Lost to PFS follow-up	Date of last assessment	Censored	Lost to PFS follow up

Item	Situation	Date of Progression or Censoring	Outcome	Censoring reason reported
3	Progressive disease with missing assessment date	Date of last adequate response assessment	Censored	Date of disease progression missing
4	Progressive disease or death after two or more missing response assessments	Date of last response assessment before missed visits	Censored	Progressive disease or death after two or more missing response assessments
5	Start of new anti-MF treatment	Date of last assessment prior to the new anti-MF treatment	Censored	Start of new anti-MF treatment

6.6.4.9.2 Descriptive Analyses

- An overall summary of number and percentage of patients whose observed PFS is ≥12, ≥18, ≥24, ≥36 months and ≥48 months will be presented for each arm.
- The number and percentage of patients with PFS event along with causes of PFS events and number of patients with censoring along with reasons for censoring will be summarized for each arm.
- The distribution of PFS will be estimated for each arm using the Kaplan-Meier (K-M) method. The median PFS time along with 95% confidence intervals will be presented (Brookmeyer and Crowley 1982). The 25th and 75th percentiles will be estimated as well, along with their 95% CI.
- The PFS rate at specific time points (12 months, 18 months, 24 months, 36 months, and 48 months), and the associated 95% CIs (Greenwood formula) will be summarized for each arm.
- The follow-up time for PFS will be calculated using reverse Kaplan-Meier methodology.

A plot of the Kaplan-Meier curve for PFS will be provided for each arm.

6.6.4.9.3 Inferential Analyses

The stratified log-rank test using the stratification factors will be used to compare PFS between the 2 treatment groups. The stratified Cox model will be used to estimate the hazard ratio and its 2-sideded 95% CIs for the treatment effects. The Kaplan Meier (K-M) survival curves and K-M medians (if estimable), along with their 2-sided 95% CIs will also be provided for each treatment group. The plots will also include the number of patients at risk at certain time points.

6.6.4.10 Overall Survival

6.6.4.10.1 Definition

OS, defined as the period from the date of randomization until the date of death from any cause. OS analyses will be based on ITT population.

OS (months) = (date of death /censoring – randomization date +1)/30.4375

Table 9: Censoring Rules for Overall Survival

Item	Situation	Date of Event/Censoring	Outcome	Censoring Reason
1	No event until data cut-off	The earlier date out of date of last contact and date of	Censored	Follow-up ongoing without event

Item	Situation	Date of Event/Censoring	Outcome	Censoring Reason
		data cut-off		
2	Death	Date of death	Event	Not Applicable
3	Lost to follow-up	Date of last contact	Censored	Lost to OS follow-up

6.6.4.10.2 Descriptive Analyses

- An overall summary of number and percentage of patients whose observed survival is ≥12, ≥24, ≥36 months and ≥48 months will be presented for each arm.
- The number and percentage of patients with OS event along with causes of death and number of patients with censoring along with reasons for censoring will be summarized for each arm.
- The distribution of OS will be estimated for each arm using the Kaplan-Meier (K-M) method. The median OS time along with 95% confidence intervals will be presented (Brookmeyer and Crowley 1982). The 25th and 75th percentiles will be estimated as well, along with their 95% CI.
- The OS rate at specific time points (12 months, 18 months, 24 months, 36 months and 48 months), and the associated 95% CIs (Greenwood formula) will be summarized for each arm.
- The follow-up time for OS will be calculated using reverse Kaplan-Meier methodology.

A plot of the Kaplan-Meier curve for OS will be provided for each arm.

6.6.4.10.3 Inferential Analyses

Treatment comparison will be performed using a stratified log-rank test, with the randomization stratification factors as reported in the IRT. OS will be tested at Primary Analysis, and at Final Analysis.

Estimates of the treatment effect will be expressed as HRs, using a stratified Cox proportional-hazards analysis randomization stratification factors as reported in the IRT. For the Primary Analysis and final analyses, 95% CIs will be estimated using Wald method.

6.6.4.11 Transformation to Acute Myeloid Leukemia (AML)

Patients will be categorized as transformation to AML if there's an increase in peripheral blood blast percentage of ≥20% that persists for at least 2 weeks or confirmed leukemic transformation per the disease status assessment.

The number and percentage of patients with transformation to blast phase (AML) for each treatment group will be tabulated for the double-blind treatment period.

A separate table will be generated to present the number and percentage of patients converted to AML in the crossover treatment period. Only patients entered the crossover cycle will be included in this summary.

A separate table will be generated to present the number and percentage of patients who developed AML related AEs by treatment group.

6.6.4.12 Duration of Splenic Response (DoSR) by Central/Local Read

Duration of the splenic response (DoSR) is defined as the time from when the criterion for splenic response is first met (i.e., a ≥ 35% reduction from baseline spleen volume) until the time at which a <35% reduction in spleen volume from baseline and also an increase of >25% from nadir as measured by MRI or CT is first documented.

DoSR will be calculated based on central read and local read separately.

Duration of the splenic response is computed in weeks as specified as below, where:

- *Date of first splenic response:* it is the first date at which a ≥35% reduction from baseline spleen volume as measured by MRI or CT is observed prior to start of new anti-cancer treatment or Progressive disease
- *Date of loss of splenic response*: for patients who had reached a ≥35% reduction from baseline spleen volume, it is the first date at which the following is observed: spleen volume as measured by MRI or CT is no longer reduced by at least 35% from baseline (i.e., spleen volume reduction <35%) and was increased by ≥ 25% from nadir (bottom).
- *Date of last adequate splenic assessment:* for patients who had reached a ≥35% reduction from baseline spleen volume, it is the date of the last MRI or CT splenic assessment at which a loss of response (i.e., spleen volume reduction <35% and 25% increase from nadir as measured by MRI or CT) is not observed. If a patient only has one MRI or CT assessment prior to discontinuation or data cut-off, then this will be considered as 'date of last adequate splenic response', too.

Table 10: Computation Rules for Duration of the Splenic Response

Item	Situation	Computation of duration of splenic response	Censoring (censored/not censored)	Censoring reason
1	Patients with a splenic response who subsequently had a loss of splenic response	(Date of loss of splenic response – date of first splenic response + 1) / 7	Not censored	
2	Patients with splenic response who discontinued from treatment prior to loss of splenic response	(Date of last splenic assessment – date of first splenic response + 1) / 7	Censored	Discontinued the study without event
3	All other patients with a splenic response	(Date of last splenic assessment - date of first splenic response + 1) / 7	Censored	Ongoing patient without lost response at cutoff
4	Event: death	Date of death	Not censored	N.A.
5	Patients receive a new anti-cancer treatment	(Date of last splenic assessment prior to date of new anti-cancer treatment - date of first splenic response + 1) / 7	Censored	New anti-cancer treatment

Patients evaluable for the duration of the splenic response are the patients in the ITT Set who had a splenic response.

Duration of splenic response will be analysed by both local and central read as follows:

- A descriptive summary of duration of splenic response will be provided for both treatment groups. For K-M, the number of patients at risk and the number of patients censored will be summarized.
- The number and percentage of patients with Splenic Response event and the number and percentage of patients with censoring along with reasons for censoring will be summarized for both treatment groups.

- The distribution of the duration of spleen response will be estimated using the Kaplan-Meier (K-M) method. For both treatment groups, K-M estimate (%) and 95% confidence limits for the K-M estimate (calculated with Greenwood's formula) will be provided at baseline and then every 12 weeks until EOS. Median, 25th and 75th percentile for survival time with 95% confidence limits will also be displayed. The confidence limits are constructed using Brookmeyer and Crowley (1982).
- Percent of SVR Responders estimate after 12 weeks, 24 weeks, 36 weeks, 48 weeks, and 60 weeks based on KM estimates and their 95% CI (using Greenwood estimates) will be provided for both treatment groups.

Kaplan-Meier (K-M) method will be presented in tables and displayed graphically.

In addition, K-M estimate of duration of follow-up of splenic response will be provided. To summarize the follow-up time based on reverse K-M method, a K-M curve will be created:

- Patients with a splenic response who subsequently had a loss of splenic response are censored with their date of loss of splenic response as the censoring date.
- Patients who are censored in the analysis of duration of splenic response above are considered as events, with the date of last adequate splenic assessment as the date of the event.

Then, Q1, median, Q3 and their 95% CI based on the KM estimates (using Brookmeyer and Crowley 1982 method) will be provided for both treatment groups.

A sensitivity of duration of spleen response analysis will also be performed employing a different definition. The alternative duration of the splenic response (DoSR) is defined as the time from when the criterion for splenic response is first met (i.e., a ≥ 35% reduction from baseline spleen volume) until the time at which a <35% reduction in spleen volume from baseline as measured by MRI or CT is first documented. The same analyses will be repeated for these sensitivity analyses.

6.6.4.13 mTSS at Week 24

The modified TSS (mTSS) is defined at the TSS score without fatigue sub-domain.

The Week n mTSS is the average of non- missing daily total symptom scores excluding fatigue over Week n. The weekly TSS will be considered missing if there are less than 4 daily total symptom scores available for that week.

Patients will be classified as responders if percentage of change from baseline mTSS ≤ -50% at Week 24 per visit window in Section 9.7.

Patients will be classified as non-responders if any of the following occurred:

1. Percentage of change from baseline mTSS > -50% at Week 24
2. Missing Week 24 mTSS after adjusted per visit window
3. Death, start of a new anti-cancer therapy or progressive disease prior to Week 24

The same analysis approach as in Section 6.6.3.2 will be provided.

6.6.4.14 Duration of TSS Response

Duration of TSS response is defined as the time from onset of TSS50 response until the time at which a <50% reduction in TSS from baseline and an increase of ≥25% from nadir.

The duration of TSS response will be measured in weeks, as follows:

Duration of TSS50 response = (TSS Response End Date – TSS50 Onset Date +1)/ 7

TSS50 Onset Date is defined as: it is the first date of the 7-days period at which a ≥ 50% reduction from baseline in TSS is observed.

TSS Response End Date is defined as one the following dates, whichever comes first:

1. Date of loss of response - the first day of the 7-days period which meeting both criteria:
   a. The average daily symptom score for a 7-day period > 50% of baseline TSS
   b. The average daily symptom score for a 7-day period exceeds 25% of the nadir TSS score
2. Date of death
3. Date of study discontinuation
4. Date of initiation of a new anti-cancer treatment

The TSS50 response calculation is described in Section 6.6.3.1. Cumulative duration of the TSS50 response is defined as the total time in response considering all episodes of response.

The censoring rule is defined as:

Table 11: Censoring Rules for Duration of the TSS50 Response

Item	Situation	Computation of duration of TSS	Censoring (censored/not censored)	Censoring reason
1	Event: Loss of response	(Date of event - date of first TSS response + 1)/7	Events	N. A
2	Event: Death	(Date of death - date of first TSS response + 1)/7	Events	N. A
3	No event (Ongoing response)	(Date of data cutoff -- date of first TSS response + 1)/7	Censored	Ongoing patient without lost response at cutoff
5	Patient discontinues the study with no event	(Last contact date - - date of first TSS response + 1)/7	Censored	Discontinued the study without event
6	Patients receive a new anti-cancer treatment	(Date of first dose of the new anti-cancer treatment - date of first TSS response + 1)/7	Censored	Initiation of new anti-cancer treatment

Duration of the TSS50 response will be summarized descriptively by treatment group for the double-blind treatment period. In addition, the number and percentage of patients without a TSS50 end date will be presented for each treatment group.

- The distribution of duration of TSS will be estimated using the Kaplan-Meier (K-M) method. For both treatment groups, K-M estimate (%) and 95% confidence limits for the K-M estimate (calculated with Greenwood's formula) will be provided at baseline and then every 12 weeks until EOS. Median, 25th and 75th percentile for survival time with 95% confidence limits will also be displayed. The confidence limits are constructed using Brookmeyer and Crowley (1982). The number and percentage of patients with duration of TSS event along with causes of death and number of patients with censoring along with reasons for censoring will be summarized for each arm.
- Kaplan-Meier (K-M) method will be presented in tables and displayed graphically.

In addition, K-M estimate of duration of follow-up of TSS will be provided. To summarize the follow-up time based on reverse K-M method, a K-M curve will be created.

A sensitivity of duration of TSS50 response analysis will also be performed employing a different definition. The alternative duration of the TSS50 response is defined as the time from when the criterion for TSS50 response is first met (i.e., a ≥ 50% reduction from baseline TSS) until the time at which a <50% reduction in TSS from baseline is first observed. The same analyses will be repeated for these sensitivity analyses.

6.6.5 Exploratory Endpoints Analysis

The following exploratory endpoints may be calculated and summarized for each treatment group for the ITT Set. No formal testing will be conducted for exploratory efficacy variables. All variables will be summarized using summary statistics. Continuous variables will be summarized using mean, standard deviation, median, Q1, Q3, minimum, and maximum values. Categorical variables will be summarized using count and percentage. Ninety-five percent confidence interval (CI) will be provided wherever appropriate. Listings for individual data will be presented.

6.6.5.1 Change from Baseline in Splenic Volume

The absolute change of the spleen volume and the percent change from baseline for the double-blind treatment period will be summarised by treatment group and overall. Analysis timepoint is provided in Table 14. Spleen volume calculation is described in Section 6.6.1.1.

The percent change from baseline in splenic volume differences between two groups at Week 12, 24, 36, 48 and 60 will also be compared using, 2 sample Kolmogorov-Smirnov test, Wilcoxon- rank-sum test and Hodges-Lehmann method to estimate the treatment difference. P-values obtained from the tests and the 95% CI obtained using Hodges-Lehmann method will be presented.

A cumulative distribution frequency plot will be provided, with vertical dotted lines at -35% on the x-axis.

6.6.5.2 RBC Transfusion Dependence at Week 24

RBC transfusion dependence at Week 24, defined as ≥ 6 units of RBC transfused during the 12-week period prior to study day 168.

The number and percentage of patients with RBC TD at Week 24 will be summarised per treatment group and overall. Percentage will be calculated based on the number of patients on the double-blind treatment period at Week 24.

6.6.5.3 Hemoglobin Response

Hemoglobin response is defined as a ≥ 1.5 g/dL average increase in hemoglobin from baseline in any 12-week mean hemoglobin concentration, as calculated over any rolling 12-week (84 days) period post-baseline in the absence of transfusions and up to the latest Hgb assessment for each patient. The response of patients with end of treatment prior to Week 12 (Day 84) will be set as non-responders. The hemoglobin baseline value will be the last observation prior to the first dose of Pelabresib with the consideration of below 14/3 - Day Rule.

<u>14/3 – Day Rule</u>

Hgb values that are collected at least 14 days after a transfusion may be used; If there is another transfusion within 3 days after the Hgb assessment, Hgb value can be still used (likely this value is low and therefore the transfusion within 3 days was required due to low Hgb). If there is a transfusion on the same day of Hgb assessment, Hgb value can be used (assumption: Hgb value is assessed prior to transfusion, therefore no transfusion impact)

[redacted]

[redacted]

[redacted]

6.6.5.8 Pharmacodynamic Effects

Below Pharmacodynamic effects will be characterized for each treatment group and overall, for the Biomarker Analysis Set.

- Post-treatment percentage changes from baseline in circulating concentrations of cytokines
- Post-treatment percentage changes from baseline in the ratio of mutant to wild type JAK2, [redacted]

6.6.5.9 Time to SVR35 response

Time to SVR35 response is defined as the time from the randomization until the first day that achieved ≥35% reduction from baseline in SVR from central read. The analysis will be based on ITT population and exclude patients without SVR35 response.

Time to SVR35 response will be summarized descriptively by treatment group for the double-blind treatment period. In addition, the number and percentage of patients without a time to SVR35 response by reasons will be presented for each treatment group.

The distribution of time to SVR35 will be estimated using the Kaplan-Meier (K-M) method. For both arms, K-M estimate (%) and 95% confidence limits for the K-M estimate (calculated with Greenwood's formula) will be provided at baseline and then every 12 weeks until EOS. Median, 25th and 75th percentile for survival time with 95% confidence limits will also be displayed. The confidence limits are constructed using Brookmeyer and Crowley (1982). The number and percentage of patients with time to SVR35 events and number and percentage of patients with censoring along with reasons for censoring will be summarized for each arm.

Kaplan-Meier (K-M) method will be presented in tables and displayed graphically.

[redacted]

[redacted]

6.7 Safety Analysis

All safety analyses will be conducted using the Safety Set. The safety endpoints for this study are as follows:

- Adverse Events
- Clinical Laboratory assessments including:
  - Hematology
  - Clinical Chemistry and Iron Studies
  - Coagulation parameters
  - Pregnancy test (WOCBP only)
  - HbA1c and serum lipidpanel
- Electrocardiograms (ECG)

- Physical examinations including spleen examination
- Vital signs

All safety endpoints except adverse events will be summarized descriptively at baseline and for each post-baseline visit. Shift from baseline will be provided, if applicable.

Data will be presented separately for the double-blind treatment period and the crossover period. Change from the crossover baseline, and shift from crossover baseline will be provided, if applicable. The crossover baseline will be defined as the last assessment before the first dose of pelabresib in the crossover period.

The following observation periods are defined for safety analyses:

- The on-treatment period starts at the time of study treatment initiation through 30 days after the last dose of study treatment.
- The crossover treatment period starts at the time of 1st dose of pelabresib after crossed over through 30 days after the last dose of cross-over study treatment.
- The overall-study observation period starts at the time of the start of the study treatment through the end of the study.

6.7.1 Subgroup Analysis of TEAE and AESIs

Subgroup analysis of TEAE and AESIs will be prepared for the following subgroups.

- Race
- Ethnicity
- Region
- Gender
- Age group
  - <65 years of age
  - ≥65 years of age
- DIPSS

6.7.2 Adverse Events

Adverse events will be coded using MedDRA Version 26.0.

A treatment-emergent adverse event (TEAE) for the double-blind treatment period is defined as an AE (classified by preferred term) that has a start date on or after the first dose of the pelabresib/placebo and before 30 days after the last dose of pelabresib/placebo or before the start of alternative (off-study) treatment for MF, whichever occurs first. If the AE has a start date before the date of first dose but increases in severity after first dose and before 30 days post last dose will be considered a TEAE as well. An AE that occurs after the administration of the first dose of open-label pelabresib treatment will be considered treatment-emergent for the crossover treatment period. However, a TEAE for the crossover treatment period that occurs within 30 days after the last dose of pelabresib/placebo will be considered treatment- emergent for the double-blind treatment period as well.

If more than one AE with the same preferred term is reported before the date of the first dose of study assigned treatment, then the AE with the highest severity will be used as the benchmark for comparison to the AEs occurring after the start of study assigned treatment under the preferred term.

AEs that occur from the time of informed consent form (ICF) signature to first dose will be collected but not evaluated in the safety analyses. They will be listed as pretreatment adverse events.

If an adverse event date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the adverse event as pretreatment or treatment emergent. The algorithm for imputing date/time of onset will be conservative and will classify an adverse event as treatment emergent unless there is definitive information to determine it is pretreatment. Handling of missing AE start/end dates are detailed in Section 9.1.

If more than one AE occurs with the same preferred term for the same subject, then the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to study treatment.

An overall summary of AEs will be tabulated for each treatment period (double-blind and crossover), including the number and percentage of patients with any AEs and TEAEs, any serious adverse events (SAEs), any severe AEs, AEs related to study drug, AEs of Grade 3 or higher, AEs causing discontinuation of study treatment, AEs leading to withdrawals from study, fatal AEs and AEs of special interest (AESI) as well as the total number of events for each category.

TEAEs will be tabulated according to the MedDRA by system organ class and preferred term and will include the following categories:

- TEAEs
- TEAEs related to study drug
- TEAEs of Grade 3 or higher
- TEAEs related to study drug that are Grade 3 or higher in severity
- TEAEs resulting in study drug discontinuation
- TEAEs resulting in study drug dose modification
- TEAEs resulting in study drug dose interruption
- Treatment-emergent AEs resulting in death
- Treatment-emergent SAEs
- Treatment-emergent SAEs related to study drug
- Treatment-emergent SAEs resulting in study drug discontinuation
- The most commonly reported treatment-emergent AEs
- The most commonly reported treatment-emergent SAEs
- Non-serious treatment-emergent AEs (at least 5% incidence in any arm)

All AEs will be presented in a listing. Additional data listings will be presented for SAEs, AEs causing discontinuation of study treatment or withdrawal from the study, AEs leading to dose reduction or interruption, fatal AEs, AESIs and a listing of all adverse events that occurred after discontinuation of study drug.

6.7.3 Adverse Events of Special Interest (AESIs)

AESIs for this trial include:

- Treatment discontinuation syndrome, which includes TEAEs with onset date within 30 days after the date of pelabresib/placebo interruption or discontinuation. This consists of the exacerbation of MF

symptoms following interruption or discontinuation of study treatment: fever, respiratory distress, hypotension, disseminated intravascular coagulation (DIC) or multi-organ failure.

- Acute respiratory distress syndrome (ARDS) including below PTs
  - Acute respiratory distress
  - Acute respiratory distress syndrome
  - Respiratory distress
  - Respiratory failure
  - Acute respiratory failure
  - Severe acute respiratory failure.

6.7.4 Clinical Laboratory Data

Descriptive statistics for clinical laboratory values, changes from baseline at each assessment time point, and shift tables from baseline to each visit for quantitative variables will be presented for the following clinical laboratory variables. All laboratory data will also be listed.

Hematology	erythrocyte count, leukocyte count, hemoglobin, hematocrit, mean corpuscular hemoglobin, mean corpuscular volume, platelet count, neutrophils, bands/stabs, lymphocytes, monocytes, basophils, eosinophils, blasts, reticulocytes
Biochemistry	urea nitrogen, creatinine, calcium, potassium, sodium, bicarbonate, glucose, urate, aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, bilirubin, lactate dehydrogenase, phosphorus, erythropoietin, C-reactive protein, iron, iron binding capacity, ferritin, and transferrin saturation
Coagulation	prothrombin time (PT), activated partial thromboplastin time (aPTT)
Other tests	urine/serum pregnancy test
	HbA1c and serum lipid panel including total cholesterol, cholesterol low density lipoprotein (LDL), cholesterol high density lipoprotein (HDL), triglycerides

Highly sensitive urine/serum beta human chorionic gonadotropin (β-hCG) pregnancy test in female patients of childbearing potential. Pregnancy test results will only be listed.

Boxplots will be produced for the following sets of parameters by visit:

- Hematology parameters: Hemoglobin, Lymphocytes, Neutrophils, Platelets, Leukocytes.
- Chemistry parameters: Alkaline Phosphatase, Alkaline Aminotransferase, Aspartate Aminotransferase, Bilirubin, Creatinine, Glucose, Potassium.
- Coagulation parameters: Activated Partial Thromboplastin Time (sec), Prothrombin Time (sec).

Patient plots of laboratory parameters of interest by visit will be produced. Only patients with ≥5x ULN or >3x baseline value of any parameters of interest will be included in the display. Parameters of interest: Aspartate Aminotransferase, Alanine Aminotransferase, Alkaline Phosphatase, Total Bilirubin, Creatinine, Albumin, Prothrombin Time.

An eDISH plot of Maximum Total Bilirubin vs. Maximum Alanine Aminotransferase will be produced. The x-axis will show Maximum Alanine Aminotransferase and will be visualized in logarithmic scale; the y-axis will show Maximum Total Bilirubin.

An eDISH plot of Maximum Total Bilirubin vs. Maximum Aspartate Aminotransferase will be produced. The x-axis will show Maximum Aspartate Aminotransferase and will be visualized in logarithmic scale; the y-axis will show Maximum Total Bilirubin.

6.7.5 Electrocardiograms (ECG)

A single 12-lead ECG will be obtained at the time points noted in Table 1 at least 1 hour after dosing. Additional unscheduled ECGs should be performed per investigator judgement (e.g., in the event of ≥ Grade 3 plasma potassium increase or hyperkalemia). The Investigator will evaluate the clinical significance of the ECGs. Clinically significant abnormal findings will be reported as AEs.

ECG parameters include PR Interval (msec), QT Interval (msec), QTcF (msec), QRS duration (msec), and the result of ECG (i.e., normal, abnormal not clinically significant, abnormal clinically significant).

Descriptive statistics of each continuous parameter will be summarized by visit, including observed values and change from baseline.

Shift tables will be produced at each visit, presenting the number and percentage of patients in each bivariate category (i.e., baseline vs. each post-baseline assessment) with regards to the result of ECG (Normal, Abnormal NCS, Abnormal CS).

Number and percentage of patients with QTcF 450-480, 480-500 msec, and QTcF >500 msec will be tabulated.

Number and percentage of patients with QTcF increase from baseline >30 ms and increase from baseline >60 ms will also be tabulated.

Number and percentage of patients with PR duration (ms) increase > 25% from baseline and PR duration >20 will be tabulated.

Number and percentage of patients with QRS duration (ms) increase > 25% from baseline and QRS duration >110 will be tabulated.

Number and percentage of patients with heart rate ≤ 50 and decrease from baseline of ≥ 25%; and ≥ 120 and increase from baseline of ≥ 25% will be tabulated separately. Tables will be generated separately for the double-blind treatment period and the crossover treatment period.

Data for all timepoints will also be listed.

6.7.6 Physical Examinations

Physical examinations will be performed at timepoints specified in Schedule of Activities in Table 1. A complete physical examination will be performed including general appearance, HEENT, neck, cardiovascular, thorax/lungs, breasts, abdomen, genitourinary, musculoskeletal, lymph nodes, skin, neurological and mental status examination. Abdominal examination will include splenomegaly and hepatomegaly. The edge of the spleen shall be determined by palpation, and measured in centimeters, using a soft ruler from the costal margin to the point of greatest splenic protrusion. Targeted physical exams may be focused on areas of known disease and potential areas of MF involvement. A targeted physical examination must include weight and examination of the abdomen to assess the spleen length by palpation.

Abnormalities identified at the screening visit will be documented in the subject's source documents and on the medical history CRF. Pre-existing conditions diagnosed through assessments and examinations at the screening visit or during the screening period are not AEs, but are recorded as medical history. If any clinically significant abnormal findings are reported after informed consent or if any pre-existing conditions worsen during the study, these must be recorded as AEs. Changes after the screening visit will be captured as AEs on the AE CRF page, as deemed clinically relevant by the investigator.

Descriptive statistics for spleen length by palpation at each post-baseline visit will be presented. Tables will be generated separately for the double-blind treatment period and the crossover treatment period.

Physical examinations performed for all timepoints will be listed.

6.7.7 Vital Signs

Descriptive statistics for vital signs (e.g., systolic and diastolic blood pressure, pulse rate, temperature, weight and respiratory rate) and their changes from baseline at each post-baseline visit will be presented by treatment group for the double-blind treatment period. Baseline will be defined as the last assessment before the first dose of study treatment.

Descriptive statistics for vital signs and their changes from crossover baseline at each post- crossover baseline visit will be presented the crossover period. Crossover baseline will be defined as the last assessment before the first dose of pelabresib in the crossover period.

Body temperature will be reported in Celsius. Weight will be reported in Kilograms.

Listings of all vital sign parameters will be provided.

6.7.8 Time to Onset of Any Grade/ $\geq$ Grade 3 TEAE

These analyses may be performed for particular AEs as well as for some laboratory abnormalities, as appropriate, for example either ALT or AST, TBIL, FPG.

Time to onset of any grade/> grade 3 event will be summarized using the Kaplan-Meier method. Median time to onset and 95% CI will be provided. In addition, Kaplan-Meier plots will be generated. If any designation other than grade 2 is required, this should be mentioned in study RAPs.

Time to onset of any grade/> grade 3 event is defined as the time from the start of treatment to the start date of the first incidence of an event of any grade/> grade 3 i.e., time in days is calculated as (start date of first occurrence of the event) – (date of first dose of pelabresib/placebo) +1.

A patient will be censored if:

- The patient discontinues from both components of study treatment without experiencing any grade/ > grade 3 event during the on-treatment period.
- The patient dies without experiencing any grade/> grade 3 event
- The patient receives a new anti-cancer therapy before experiencing any grade/> grade 3 event.
- The patient is still ongoing at the analysis cut-off date without experiencing any grade/> grade 3 event

The censoring date will be the earliest of the following dates: end of last dose of pelabresib/placebo + 30 days, new anti-cancer therapy start date, death date, or last contact date.

For lab events the censoring date will be the date of the last scheduled/unscheduled lab assessment with a non-missing value for the lab parameter of interest that was done prior to the time of occurrence of the censoring reason.

6.8 Treatment Exposure

All analyses under this section will be conducted for the Safety Set.

Unless otherwise specified, study treatment of pelabresib/placebo will be administered QD for 14 consecutive days followed by a 7-day break, which is considered 1 cycle of treatment (21 days). Ruxolitinib will be administered BID for all 21 days within each cycle.

The measures of exposure will be summarized using descriptive statistics (n, mean, standard deviation, minimum, median, and maximum) for each individual study treatment component (pelabresib/placebo and ruxolitinib). The following exposure measures will be provided:

- Total Cumulative Dose Received
    - Defined as total dose in mg received from first dose date up to and including last dose date based on drug accountability.
- Total Cumulative Dose Prescribed
    - Defined as the sum of the prescribed total daily doses in mg for each cycle up to the last dose date.
    - Any dose adjustment and interruptions will be counted into the total cumulative dose prescribed
- Average Daily Dose
    - Defined as the total cumulative dose received/total days of drug exposure
- Duration of Treatment (weeks) (to be calculated using KM analysis)
    - Defined as (the date of last dose – the date of first dose +1)/7, within both analysis periods combined.
- Categories of treatment duration: ≥12 weeks, ≥24 weeks, ≥36 weeks, and ≥48 weeks.
- Number of Cycles Received
    - Defined as the number of cycles where a patient receives study treatment
- Dose Intensity Received
    - Defined as [the total cumulative dose received/duration of treatment (weeks)]
- Dose Intensity Prescribed
    - Defined as [the total cumulative dose prescribed/duration of treatment (weeks)]

For patients who have not discontinued treatment permanently, the most recent completed cycle last day will be used in place of "date of last dose".

Pelabresib/placebo drug accountability will be obtained in the form of a data transfer as provided by Suvoda's Interactive Response Technology (IRT) system. Ruxolitinib drug accountability will be obtained from the "Ruxolitinib Drug Accountability" eCRF.

The number and percentages of patients who discontinued pelabresib/placebo or ruxolitinib permanently, according to entries on the eCRFs, and the number and percentages of patients who discontinued pelabresib or ruxolitinib permanently because of an AE will be summarized by treatment group.

Details of study drug exposure to pelabresib/placebo and ruxolitinib will be presented in data listings.

6.9 Measurement of Treatment Compliance

For each patient, compliance, reflecting the patient's intake of medications as prescribed, will be calculated for each individual study treatment component (pelabresib/placebo and ruxolitinib):

- Treatment Compliance = [Total Cumulative Dose Received/Total Cumulative Dose Prescribed] x 100%

Total Cumulative Dose Received will be based on drug accountability. Only completed cycles prior to the cutoff date will be included in the compliance calculation.

The Total Cumulative Dose Prescribed is defined as the number of doses that the patient would have received up to the date of the patient's withdrawal or termination from the study. Any reported dose adjustment and interruptions will be counted in the expected number of doses calculation.

Treatment compliance will be summarized using descriptive statistics (n, mean, standard deviation, minimum, median, and maximum) by individual study treatment component (pelabresib/placebo and ruxolitinib) for the Safety Set. Overall study period compliance and compliance for the first 24 weeks will be provided.

The following summaries for dose adjustment and dose interruption will be provided:

Dose adjustment:

- The number and percentage of patients with at least one dose adjustment (increased, decreased, frequency change), and the cycle of first dose adjustment
- The number and percentage of patients who have dose adjustment due to adverse events

Dose interruption:

- The number and percentage of patients with at least one interruption, total number of interruptions, and reason for interruption.

Pelabresib/placebo and ruxolitinib dose adjustment and interruption will be listed and summarized by cycle, by overall up to Week 24, and by overall within the double-blind analysis period and overall study period.

6.10 Pharmacokinetic Analysis

Plasma concentration data of pelabresib and ruxolitinib, and plasma concentration data of metabolites M542/M544 will be summarized for all patients who receive at least 1 dose of study drug and who have at least 1 quantifiable concentration, regardless of their inclusion in the PK population. PK concentration values will be listed for each individual by study drug, treatment arm, cycle, day, and dose, and the following descriptive statistics will be provided: N (number of patients with non-missing data), mean (arithmetic mean), standard deviation (SD), arithmetic percent coefficient of variation (CV%), minimum, median, and maximum.

Descriptive and graphical methods will be used to explore the effects, if any, of pelabresib on the PK of ruxolitinib.

Population PK assessment to characterize the PK of pelabresib will be described in detail in a separate analysis plan.

6.11 Biomarker Analysis

Biomarker endpoints included plasma cytokine levels and blood mutation profiling before and after start of treatment (ratio of mutant to wild type MF-relevant alleles).

The following listings will be generated (Biomarker Analysis Set):

- Mutation Profiles: this listing will report individual patients' data on a selection of blood allelic burden genes

- Cytokine Profiles: this listing will report individual patients' data on plasma cytokine levels and will include absolute and percent change from baseline

6.12 Patient Reported Outcomes (PROs)

The change over time in patient rating of severity of MF symptoms, global impression of change, and quality of life will be analysed in secondary endpoints and exploratory endpoints using the following PRO measures, respectively:

- MFSAF version v4.0
- EQ-5D-5L
- PGIC

Detailed analyses based on MFSAF are in Section 6.6.2, Section 6.6.3, Section 6.6.4.1, Section 6.6.4.4 and Section 6.6.4.13. Refer to Section 6.6.5.5 for EQ-5D-5L and Section 6.6.4.8 for PGIC.

6.13 Interim Analyses

No formal interim analysis for efficacy will be performed.

7 REFERENCES

Arber DA., Orazi A., Hasserjian R., et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016; 127: 2391-405.

Blum KA., Abramson J., Maris M., et al. A phase I study of CPI-0610, a bromodomain and extra terminal protein (BET) inhibitor in patients with relapsed or refractory lymphoma. Annals of Oncology. 2018; 29.

Brookmeyer, R., & Crowley, J. (1982). A confidence interval for the median survival time. Biometrics, 29-41.

Carpenter, J. R., Roger, J. H. , & Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation. Journal of Biopharmaceutical Statistics, 23(3), 1352–1371.

Ciurea SO., Merchant D., Mahmud N., et al. Pivotal contributions of megakaryocytes to the biology of idiopathic myelofibrosis. Blood. 2007; 110: 986-93.

Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates; 1988.

Gerds AT. Myeloproliferative Neoplasms. Cleveland Clinic Center for Continuing Education. 2016.

Harrison CN, Vannucchi AM, and Platzbecker U. Phase 3 randomized trial of momelotinib (MMB) versus best available therapy (BAT) in patients with myelofibrosis (MF) previously treated with ruxolitinib (RUX). J Clin Oncol. 2017a; 35: 7001.

Harrison CN., Patriarca Andrea, Mascarenhas John, et al. Preliminary Report of MANIFEST, a Phase 2 Study of CPI-0610, a Bromodomain and Extraterminal Domain Inhibitor (BETi), in Combination with Ruxolitinib, in JAK Inhibitor (JAKi) Treatment Naïve Myelofibrosis Patients. Blood. 2019; 134: 4164-64.

Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Cambridge, MA, USA: Academic Press; 1985

Kleppe M, Koche R, Zou L, et al. Dual targeting of oncogenic activation and inflammatory signaling increases therapeutic efficacy in myeloproliferative neoplasms. Cancer Cell. 2018; 33: 785-87.

Kuter DJ., Bain B., Mufti G., et al. Bone marrow fibrosis: pathophysiology and clinical significance of increased bone marrow stromal fibres. Br J Haematol. 2007; 139: 351-62.

Mascarenhas J, Kremyanskaya M, Hoffman R, et al. MANIFEST, a Phase 2 Study of CPI-0610, a Bromodomain and Extraterminal Domain Inhibitor (BETi), As Monotherapy or "Add-on" to Ruxolitinib, in Patients with Refractory or Intolerant Advanced Myelofibrosis. Blood. 2019; 134: 670.

Mehrotra D. and Railkar R. (2000). Minimum risk weights for comparing treatments in stratified binomial trials. Statistics in Medicine, 19(6):811-825

Mertz J, Blum K, Younes A, et al. Pharmacodynamic assessment in whole blood for the BET bromodomain inhibitor CPI-0610 of target engagement in patients with progressive lymphoma. AACR Annual Meeting. 2018.

Mesa RA, Kiladjian JJ, Catalano JV, et al. SIMPLIFY-1: A phase III randomized trial of momelotinib versus ruxolitinib in Janus Kinase Inhibitor-Naive patients with myelofibrosis. J Clin Oncol. 2017; 35: 3844-50.

Naymagon L., and Mascarenhas J. Myelofibrosis-Related Anemia: Current and Emerging Therapeutic Strategies. Hemasphere. 2017; 1: e1.

NCCN. Guidelines for Patients - Myeloproliferative Neoplasms. 2019: 1-70.

Newberry KJ, Patel K, Masarova L, et al. Clonal evolution and outcomes in myelofibrosis after ruxolitinib discontinuation. Blood. 2017; 130: 1125-31

Nicodeme E, Jeffrey KL, Schaefer U, et al. Suppression of inflammation by a synthetic histone mimic. Nature. 2010; 468: 1119-23.

Palandri F., Palumbo G. A., Elli E. M., et al. Ruxolitinib discontinuation syndrome: incidence, risk factors, and management in 251 patients with myelofibrosis. Blood Cancer J. 2021; 11: 4.

Pikman Y., Lee BH., Mercher T., et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med. 2006; 3: e270.

Shanavas M., Popat U., Michaelis L. C., et al. Outcomes of Allogeneic Hematopoietic Cell Transplantation in Patients with Myelofibrosis with Prior Exposure to Janus Kinase 1/2 Inhibitors. Biol Blood Marrow Transplant. 2016; 22: 432-40.

Talpaz M., Erickson-Viitanen S., Hou K., et al. Evaluation of an alternative ruxolitinib dosing regimen in patients with myelofibrosis: an open-label phase 2 study. J Hematol Oncol. 2018; 11: 101.

Tefferi A, Lasho TL, Jimma T, et al. One thousand patients with primary myelofibrosis: the mayo clinic experience. Mayo Clin Proc. 2012; 87: 25-33.

Tefferi A., and Barbui T. Polycythemia vera and essential thrombocythemia: 2019 update on diagnosis, risk-stratification and management. Am J Hematol. 2019; 94: 133-43.

Tefferi A., Cervantes F., Mesa R., et al. Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. Blood. 2013; 122: 1395-8.

Tefferi A., Vaidya R., Caramazza D., et al. Circulating interleukin (IL)-8, IL-2R, IL-12, and IL- 15 levels are independently prognostic in primary myelofibrosis: a comprehensive cytokine profiling study. J Clin Oncol. 2011; 29: 1356-63.

Thiele J., Kvasnicka H. M., Facchetti F., et al. European consensus on grading bone marrow fibrosis and assessment of cellularity. Haematologica. 2005; 90: 1128-32.

Vannucchi AM. Management of myelofibrosis. Hematology Am Soc Hematol Educ Program. 2011; 2011: 222-30.

Verstovsek S., Mesa R. A., Gotlib J., et al. A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. N Engl J Med. 2012; 366: 799-807.

White Ian R., Joseph Royes, Best Nicky. A causal modelling framework for reference-based imputation and tipping point analysis in clinical trials with quantitative outcome. J Biopharmaceut Stat 2020; 30(2): 334-350.

Zimmerman HJ. Drug-induced liver disease. Hepatotoxicity: the adverse effects of drugs and other chemicals on the liver. 1st ed. New York: Appleton-Century-Crofts; 1978. p. 351–3.

8 CHANGES TO PROTOCOL PLANNED ANALYSES

None. The statistical analysis plan has no changes to protocol amendment 5, version 6.0.

9 APPENDIX

9.1 Handling of Missing Dates/Months/Years for Adverse Events

The following rules are applied for handling of missing or partial dates for AE:

- If the start date of an AE is partial or missing and the end date of the AE occurs earlier than the start of treatment, then the AE is not considered as treatment-emergent.
- If the start date of an AE is partial or missing and the AE end date is partial but it is shown that it occurs before the start of treatment, then the AE is not considered as treatment-emergent.
- In all other cases of partial/missing dates of an AE, the AE will be considered as treatment-emergent. Hence, no imputation will be performed, and a worst-case scenario will be applied when defining TEAE.

9.2 Handling of Missing Dates/Months/Years for Prior/Concomitant Medication

The following rules are applied for missing/partial dates of prior and concomitant medications:

- If the start day of a medication is missing, and the start month is the same as the month of start of study drug, then the medication will be counted as both prior and concomitant, unless the end date of the medication is before the start of treatment. In this case, the medication will be counted as prior medication only.
- If the start date of a medication is before the start of study treatment but the end date of the medication is missing, then the medication is considered as concomitant.
- If the start date of a medication is before the start of study treatment and the end date of the medication is partial, the medication is considered as prior only if it is obvious that the year or month information is before the start of treatment. In all other cases, the medication is considered as concomitant.
- If start day and month are missing but the start year of a medication is the same as the year of start of treatment, then the medication is considered as both prior and concomitant, unless the end date is before start of treatment.

9.3 Handling of Missing/incomplete dates for data other than Adverse Events or prior/concomitant medications

This section describes some general principles to be followed in the case of missing or incomplete dates other than adverse events or prior/concomitant medications. The dates that are missing or incomplete are derived as follows:

- Dates are split in 3 parts: year, month and day. Year is the top-level, month is medium level and day is low level.
- If a part is missing, all other parts of a lower level are considered to be missing. This means that a DDMMMYYYY date ‘21---2021’ is considered as ‘-----2021’.

Missing parts for specific dates are changed into acceptable non-missing values as described in Table 12. In the following, ‘lower limit’ and ‘upper limit’ refer to the minimum or maximum, respectively, of a possible date. For example, if the day is missing, the lower limit is the first day of the given month and the upper limit is the last day of the given month. If the day and month are missing, the lower limit refers to the first day of the given year and the upper limit to the last day of the given year. In case more than one date is missing per patient, imputation will be performed as the order specified in Table 12.

Table 12: Handling of Missing and Incomplete Dates

Type of Date	Date is incomplete	Date is missing
Records of medical history - start date	- Day is missing: Replaced by the lower limit. - Day and month are missing: no replacement.	No replacement
Records of medical history - stop date	- Day is missing: Replaced by the upper limit. - Day and month are missing: no replacement.	No replacement
Date of measurements: Laboratory assessments, Vital signs, ECG.	- Day is missing: Replaced by the lower limit. - Day and month are missing: no replacement.	No replacement
Records of subsequent anticancer therapy	- Day is missing: Replaced by the lower limit. - Day and month are missing: no replacement.	No replacement
End of treatment date	- Day is missing: Replaced by the upper limit.	No replacement
Death date	Use the latest date among: - Lower limit - Any date record as collected in the eCRF	- Use the latest date among any date record as collected in the eCRF

9.4 Handling of Missing/Incomplete Dates for Date of Initial MF Diagnosis

The table below reports the rules to apply for partial dates of initial MF diagnosis.

Table 13: Handling of Incomplete Dates of Initial MF Diagnosis

Day	Month	Year	Treatment start	Imputation rule
Missing	Non-missing	Non-missing	Month and year for start of treatment ≥ month and year of initial MF diagnosis	Replace missing day with 1st of initial diagnosis month.
Missing	Missing	Non-missing	Year of start of treatment ≥ year of initial MF diagnosis	Replace missing day and month with 1st Jan of initial diagnosis year

In all other cases, no imputation will be performed.

9.5 Visit Windows for Spleen Volume Response (SVR)

Table 14 below illustrates the visit windows and the target study day for splenic volume data that are analyzed by visit.

Table 14: Analysis Visit Window Definition for SVR

Analysis Timepoint*	Target Study Day per Protocol	Analysis Visit Window (in days)	
		Lower Bound	Upper Bound
Baseline	Last non-missing before randomization	NA	Day 1
Week 12	Day 84	Day 64 (9*7+1)	Day 105 (15*7)
Week 24	Day 168	Day 148 (21*7+1)	Day 189 (27*7)
Week 12*k (k=3, 4, 5, …)	7*12*k	Day [(12*k-3)*7+1]	Day [(12*k+3)*7]

* Analysis Timepoint for analysis purpose is used to report data in tables, figures and listings.

SVR data collected on all scheduled or unscheduled visits will be mapped to an appropriate analysis visit window using the following rules:

1. If no numerical record is available within a visit window, the spleen volume will be considered as missing for the visit, and imputation rules will be applied under sensitivity analysis.
2. If there is only one numerical record in a visit window, this will be used.
3. If more than 1 numerical record is available within the same visit window, the record closest to the target study day will be used. If there are multiple records within the same distance from the target study day, the latest record will be used.
4. Upper Bound will be set as the date of initiation of new anti-cancer therapy if patient has started a new anti-cancer therapy during the analysis visit window.

9.6 Study Weeks for TSS

TSS for a given week is calculated as the average of non-missing daily total symptom scores (using the MFSAF v4.0) over the week under scrutiny as depicted in Table 15 below.

Daily TSS score is considered missing if there are less than 4 out of 7 questions answered.

Table 15: Study Week Definition for TSS

Study Week	TSS Assessment Day Relative to Randomization Date
Week -k (k=3, 4, 5, …)	Day (-k*7) to Day (-k*7+6)
Week -2	Day -14 to Day -8
Week -1 (Baseline Week)	Day -7 to Day -1
Week 1	Day 1 to Day 7
Week 2	Day 8 to Day 14
Week k (k=3, 4, 5, …, 23)	Day (k*7-6) to Day (k*7)

Week 24	Day 162 to Day 168
Week k (k=25, 26, 27, …)	Day (k*7-6) to Day (k*7)

On the day of randomization, if the MFSAF is completed prior to randomization then it will be treated as Day -1 of the baseline week, if the MFSAF is completed after randomization then it will be treated as Day 1 of Week 1.

9.7 Visit Windows for TSS and PGIC

Table 16 below illustrates the visit windows and the target study day for TSS data and the PGIC data.

Table 16: Analysis Visit Windows Definition for TSS and PGIC

Analysis Timepoint	Target Study Day	Analysis Visit Window	
		Lower Bound	Upper Bound
Week 12	Day 78 - 84	Day 64 (9*7+1)	Day 105 (15*7)
Week 24	Day 162 - 168	Day 148 (21*7+1)	Day 189 (27*7)
Week 12*k (k=3, 4, 5, …)	Day 7*12*k-6 to Day 7*12*k	Day [(12*k-3)*7+1]	Day [(12*k+3)*7]

For TSS:

For TSS data that are summarized or analyzed by Analysis Timepoint, the weekly averaged TSS (as calculated per Section 9.6) is mapped to an appropriate analysis timepoint, using the following rules:

1. If no numerical record is available within a visit window, the TSS value will be considered as missing for the analysis timepoint.
2. If only one weekly average TSS value is available for a given analysis timepoint visit window, this will be used. If more than one weekly average TSS are available within the same visit window, only one will be picked per prioritizing rules. Every consecutive 7-days period will be checked within the analysis visit window, not just the study week in Table 15.
3. Prioritizing rules if case of multiple weekly averaged TSS available: Closest to study day > number of assessments available for a given 7-days > Worst case scenario (largest TSS) > Earliest
4. Upper Bound will be set as the date of initiation of new anti-cancer therapy if patient has started a new anti-cancer therapy during the analysis visit window.

For PGIC:

1. If no numerical record is available within a visit window, the PGIC value will be considered as missing for the analysis timepoint.
2. If only PGIC value is available for a given visit window, this will be used. If more than one PGIC are available within the same visit window, only one will be picked per prioritizing rules.
3. Prioritizing rules if case of multiple PGIC available: Closest to study day (Day 12*k for Week k) > Worst case scenario (largest PGIC score) > Earliest Day

9.8 Visit Windows for Bone Marrow Fibrosis Grade

Table 17 below illustrates the visit windows and the target study day for the analysis of bone marrow fibrosis grade data.

Table 17: Analysis Visit Window Definition for Bone Marrow Fibrosis

Analysis Timepoint	Target Study Day	Analysis Visit Window	
		Lower Bound	Upper Bound
Baseline	-1	NA	Day -1 (the day prior to Day 1)
Week 24	168	Day 85 (12*7+1)	Day 252 (36*7)
Week 48	336	Day 253 (36*7+1)	Day 420 (60*7)
Week 24*k (k=3, 4, 5…)	7*24*k	Day [(24*k-12) *7+1]	Day [(24*k+12) *7]

Data collected on all scheduled and unscheduled visits will be mapped to an appropriate analysis visit window using the following rules:

1. If no numerical record is available within a visit window, the bone marrow fibrosis grade will be considered missing for the visit.
2. If there is an event of initiation of new anti-cancer therapy, progression disease, or crossover to open-label treatment during an analysis window, only assessments prior to the event will be considered.
3. If more than 1 numerical record is available within the same visit window, the record closest to the target study day will be used; if multiple records within the same distance from the target study day are available, the latest record will be used.

For central read data:

Bone marrow fibrosis grade central read data will be summarized or analyzed by nominal visit, because the date of sample collection is not available from central read.

Bone marrow fibrosis grade at Week 24 from central read is defined as the last non-missing post-baseline result from central read on/prior to Week 24.

Bone marrow fibrosis grade central read and local read will be matched by the nominal visit name.

9.9 Repeated or Unscheduled Assessments of Safety Parameters

Repeated or unscheduled assessments of safety parameters will be handled as follows:

If a patient has repeated assessments before the start of study treatment, then the results from the final assessment made prior to the start of study treatment will be used as baseline. If end of study assessments are repeated or unscheduled, the last post-baseline assessment will be used as the end of study assessment for generating descriptive statistics. If a patient has repeated assessments between the start of study treatment in the treatment period and the end of study visit, the assessments of unscheduled visits will be excluded in the table summary. However, all assessments will be presented in the data listings.