Official Title:	**A Randomized, Single-Blind, Multicenter Phase 2 Study to Evaluate the Activity of 2 Dose Levels of Imetelstat in Subjects With Intermediate-2 or High-Risk Myelofibrosis (MF) Relapsed/Refractory to Janus Kinase (JAK) Inhibitor**
NCT Number:	NCT02426086
Document Date:	Statistical Analysis Plan: 14 May 2018

Janssen Research & Development*

Statistical Analysis Plan

A Randomized, Single-Blind, Multicenter Phase 2 Study to Evaluate the Activity of 2 Dose Levels of Imetelstat in Subjects with Intermediate-2 or High-Risk Myelofibrosis (MF) Relapsed/Refractory to Janus Kinase (JAK) Inhibitor

Protocol 63935937MYF2001; Phase 2

JNJ-63935937 (imetelstat)

* Janssen Research & Development is a global organization that operates through different legal entities in various countries. Therefore, the legal entity acting as the sponsor for Janssen Research & Development studies may vary, such as, but not limited to Janssen Biotech, Inc.; Janssen Products, LP; Janssen Biologics, BV; Janssen-Cilag International NV; Janssen, Inc; Janssen Infectious Diseases BVBA; Janssen Sciences Ireland UC; or Janssen Research & Development, LLC. The term "sponsor" is used throughout the protocol to represent these various legal entities; the sponsor is identified on the Contact Information page that accompanies the protocol.

Issue/Report Date: 14 May 2018
Department: Biometrics and Reporting
Document No.: EDMS-ERI-162883916, 1.0

TABLE OF CONTENTS

ABBREVIATIONS

aPTT	activated partial thromboplastin time
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BM	bone marrow
CI	clinical improvement
CR	complete remission
DMC	Data Monitoring Committee
DNA	deoxyribonucleic acid
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EQ-5D-5L	EuroQol-EQ-5D-5L
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GGT	gamma-glutamyl transpeptidase
Hb	hemoglobin
HI-E	hematologic improvement-erythroid
HIV	human immunodeficiency virus
HR	heart rate
hTERT	human telomerase reverse transcriptase
IB	Investigator's Brochure
ICF	informed consent form
ICH	International Conference on Harmonisation
IEC	Independent Ethics Committee
INR	international normalized ratio
IPSS	International Prognostic Scoring System
IRB	Institutional Review Board
IRR	infusion related reaction
ITT	intent to treat
IV	intravenous
IWG	International Working Group
IWG 2006	International Working Group Response Criteria [in Myelodysplasia] 2006
IWG-MRT	International Working Group – Myeloproliferative Neoplasms Research and Treatment
IWRS	interactive web response system
LFT	liver function tests
MDS	myelodysplastic syndrome(s)
MedDRA	Medical Dictionary for Regulatory Activities
MF	Myelofibrosis
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
OS	overall survival
PD	progressive disease / disease progression
PE	physical exam
PK	pharmacokinetic(s)
PR	partial remission
pRBC	packed red blood cell
PRO	patient-reported outcome(s)
PT	prothrombin time
qRT-PCR	quantitative real-time polymerase chain reaction
RBC	red blood cell
RNA	ribonucleic acid
SIPPM	Site Investigational Product Procedures Manual
SVR	spleen volume reduction
TA	telomerase activity

TI	transfusion independence
TL	telomere length
ULN	upper limit of normal
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan contains definitions of analysis populations, derived variables and statistical methods for analysis of efficacy and safety.

1.1. Trial Objectives

The primary objective of this study was to evaluate spleen response rate and symptom response rate of 2 dose regimens of imetelstat (9.4 mg/kg and 4.7 mg/kg imetelstat given intravenously [IV] every 3 weeks) in subjects with intermediate-2 or high-risk myelofibrosis (MF) who are relapsed after or refractory to JAK inhibitor treatment.

Secondary objectives were to assess the safety of imetelstat; to assess complete remission (CR) or partial remission (PR), clinical improvement (CI), spleen response, symptoms response and anemia response per modified 2013 International Working Group – Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) criteria, duration of responses, and overall survival (OS); to evaluate the pharmacokinetics (PK) of imetelstat; to evaluate the PK/response and pharmacodynamic (PD) relationships with factors that include hemoglobin concentration, spleen size, and platelet count; to evaluate the immunogenicity of imetelstat; and to assess the effect of treatment on patient-reported outcomes (PROs).

The exploratory objectives were to assess the PK relationship with telomerase activity (TA), telomere length (TL) or human telomerase reverse transcriptase (hTERT), to characterize the baseline cytogenetic and mutational status for their association with clinical response, and to evaluate the change of cytogenetic abnormalities or mutant allele burden for assessing cytogenetic and molecular responses.

1.2. Trial Design

This was a randomized, single-blind, multicenter, Phase 2 study of 2 dosing regimens of single-agent imetelstat in subjects with intermediate-2 or high-risk MF (primary MF [PMF], post-ET--MF [PET-MF], or post-PV-MF [PPV-MF]) who were relapsed after or refractory to JAK inhibitor treatment. Eligible subjects were stratified based on a) spleen size $\geq$ 15 cm below the left costal margin (yes vs. no) and b) platelet count at study entry (platelets $\geq$75 x 10^9/L and < 150 x 10^9/L vs. $\geq$ 150 x 10^9/L). Study treatment was administered on a 21-day cycle. The main study had phases: a screening phase of up to 21 days before randomization; a treatment phase extended from randomization until

study drug discontinuation; and a posttreatment follow-up phase that began when the subject discontinued imetelstat. The posttreatment follow-up phase would continue until death, lost to follow-up, withdrawal of consent or study end, whichever occurred first. The end of study was defined as 18 months after the last subject was enrolled or when the sponsor terminated the study, whichever occurred first.

With Protocol Amendment 3, the study has been closed to further subject enrollment. Subjects in the Treatment Phase who benefited from study treatment may enter an Extension Phase and continue to receive imetelstat until there was loss of benefit or unacceptable toxicity, as determined by the investigator according to local standard of care. Subjects in the Posttreatment Follow-up Phase willenter the Extension Phase to continue follow-up for survival status. The Extension Phase will end approximately 1 year after the clinical cutoff for the final analysis of the main study.

The first interim review planned in this study occurred after the first 40 evaluable subjects (defined as the subjects who have both baseline and post baseline spleen measurement per investigator MRI, 20 per treatment arm) were followed for at least 12 weeks. Both arms would continue with enrollment until approximately 100 subjects per arm were enrolled if both of them were efficacious and had an acceptable safety profile. If only one arm was efficacious and had an acceptable safety profile, then that arm would continue enrollment and enrollment would be stopped in the other arm. In the case that both arms were not efficacious and safe, an alternative dose may be selected based on the PK/PD exposure-response, efficacy, and safety analysis and a protocol amendment would be implemented to reflect this.

Following the first interim review of data, and in line with Protocol Amendment 2, enrollment of new subjects into Arm A was suspended and enrollment into Arm B was permanently closed. Subjects in Arm B continued their current treatment or had their dose increased to 9.4 mg/kg at the investigator's discretion according to the guidance provided in Section 6 of protocol. Subjects in Arm A could have continued treatment at the investigator's discretion.

The second interim review was performed with a clinical cutoff that was 6 months after the clinical cutoff used for the first interim review. Following the second interim review, it was decided that enrollment in Arm A would be permanently closed.

The analyses specified in the SAP are considered the final analyses for the protocol, with a data cut-off of 18 months after the last subject was enrolled. Data collected in the extension phase will not be included in the analyses and will not be reported here. There will be a separate addendum with data from that phase.

The final decision for further development will be made based on the totality of data taking into consideration efficacy data, including spleen responses, symptoms, peripheral blood counts, bone marrow data, survival, safety, and PK/PD data.

1.3. Statistical Hypotheses for Primary Objectives

The primary hypothesis of this study is that the experimental treatment compared with historical data will significantly improve both spleen response rate and symptom response rate in subjects with intermediate-2 or high-risk MF relapsed/refractory to a janus kinase (JAK) Inhibitor.

Statistical Hypotheses are:

H_0: [redacted]

H_1: [redacted]

The null hypothesis is tested by the intersection-union test.[1] Each of the co-primary endpoints is tested at one sided α=0.025, and the maximum Type I error rate for testing co-primary endpoints is the same as one sided α=0.025.

1.4. Sample Size Justification

The study was planned to enroll approximately 200 subjects (approximately 100/arm) in the original protocol. As of result of decisions from the interim reviews, the study enrolled 107 subjects.

[redacted]

[redacted]

1.5. Randomization and Blinding

Subjects were centrally randomized to 2 dose regimens of imetelstat (9.4 mg/kg and 4.7 mg/kg imetelstat given intravenously [IV] every 3 weeks) in a 1:1 ratio based on a computer-generated randomization schedule prepared before the study. The randomization was stratified by a) spleen size ≥15 cm below the left costal margin by palpation (yes vs. no) and b) platelet count (≥75 x 10^9/L and <150 x 10^9/L vs. ≥ 150 x 10^9/L) and balanced by using randomly permuted blocks.

The study was originally designed as a single-blind study, in which subjects would be blinded to the treatment arms. This was to reduce potential bias during data collection. Beginning with Protocol Amendment 2, enrollment of new subjects into Arm A (9.4 mg/kg) was suspended and enrollment into Arm B (4.7 mg/kg) was permanently closed. Any subject already in the Screening Phase at the time study enrollment was suspended, and who was subsequently determined to be eligible for the study, was allowed to proceed with treatment assignment to the 9.4 mg/kg arm per investigator discretion, subject agreement, and IRB/IEC notification. In addition, subjects were no longer blinded to treatment assignment. Data review committees and their access to data is detailed in Section 2.2.2.

2. ANALYSIS PLANNED

This analysis plan describes the planned efficacy and safety analyses for Study 63935937MYF2001 (hereafter referred to as Study MYF2001).

2.1. General Analysis Definitions

2.1.1. Baseline, Treatment Phase and Posttreatment Follow Up Phase

Unless specified otherwise, the baseline value is defined as the last non-missing value collected on or before the administration of the first dose of study drug. For subjects who have been randomized but not treated with any study drug, the randomization date will be used as the reference date for baseline value calculation.

The treatment phase begins on date of Cycle 1, Day 1 and continues until study drug discontinuation.

The Posttreatment follow up phase starts from the day after study drug discontinuation until death, lost to follow-up, withdrawal of consent or study end, whichever occurs first.

2.1.2. Visit Window

The following table will be used to generate derived visit based on relative day.

Parameter	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)
Vital signs	Baseline/0	<=1	1
	4	14 to 41	28
	8	42 to 69	56
	12	70 to 97	84
	.	.	.

For Visit = week 4k, target day is 4*k*7, relative day is within window

4(k)*7-14<= relative day < 4k*7 + 14

2.1.3. Data Handling Rule and Imputation of Missing Data

In general, no missing data, other than missing dates specified below, will be imputed. For subjects who had dose escalation (from 4.7 mg/kg to 9.4 mg/kg), data up to dose escalation will be included in tables generated for CSR. Data after dose escalation (inclusive) will be listed only.

Imputation of partial missing dates will be made for AE onset date, AE resolution date, date of death, start and end dates of concomitant therapy, and date of initial diagnosis.

For AE onset and resolution date, the global standard AE imputation rules listed below will be applied.

Partial AE onset dates will be imputed as follows:

- If the onset date of an adverse event is missing day only, it will be set to:

  - First day of the month that the AE occurred, if month/year of the onset of AE is different than the month/year of the first dose date
  - The day of first dose date, if the month/year of the onset of AE is the same as month/year of the first dose date and month/year of the AE resolution date is different
  - The day of first dose date or day of AE resolution date, whichever is earliest, if month/year of the onset of AE and month/year of the first dose date and month/year of the AE resolution date are same

- If the onset date of an adverse event is missing both day and month, it will be set to the earliest of:
  - January 1 of the year of onset, as long as this date is after the first dose date
  - Month and day of the first dose date, if this date is the same year that the AE occurred
  - The AE resolution date.
- Completely missing onset dates will not be imputed.
- Partial AE resolution dates not marked as ongoing will be imputed as follows:
  - If the resolution date of an adverse event is missing day only, it will be set to the earliest of the last day of the month of occurrence of resolution or the day of the date of death, if the death occurred in that month.
  - If the resolution date of an adverse event is missing both day and month, it will be set to the earliest of December 31 of the year or the day and month of the date of death, if the death occurred in that year.
  - Completely missing resolution dates will not be imputed.

For start and end dates of concomitant therapies, the global standard AE imputation rules listed above will be applied accordingly.

For date of death and date of initial diagnosis, the following rule will be applied.

- If date is completely missing, no imputation will be made.
- If year is missing, no imputation will be made.
- If only year is present but month and day are missing, then June 30th will be used.
- If only day is missing but year and month are available, then the 15th of the month will be used.

However, the above imputations will be modified by the following rules:

- If such imputed date for initial diagnosis is on or after study consent date, then consent date -1 will be used
- If the imputed date is for a date of death and is before the last date that the subject is known to be alive, the latter date + 1 will be used.

2.1.4. Analysis Sets

2.1.4.1. Treated Analysis Set

The Treated analysis set is defined as all subjects who receive at least 1 dose of study drug. It will be used as both the primary efficacy analysis set and safety analysis set.

2.1.4.2. Intent-to-Treat Analysis Set

The Intent-to-Treat (ITT) analysis set is defined as all subjects randomized into the study and who will be classified according to their assigned treatment group. It will be used for analyses of disposition, demographic, and baseline disease characteristics.

2.1.4.3. Biomarker Analysis Set

The Biomarker analysis set is defined as all subjects who receive at least 1 dose of study drug and have at least one post baseline biomarker assessment done.

2.1.4.4. Treated for Week 12/24 Analysis Set

The Treated for Week 12/24 analysis set is defined as all subjects who receive at least 1 dose of study drug with an observed measurement collected at week 12/24.

2.1.4.5. Definition of Subgroup

Subgroup analysis will be performed for the selected variables to assess the internal consistency of the treatment benefit and/or safety. The subgroup variables and the cutoff values are subject to change if warranted to better represent the data. Additional subgroup analyses (not limited to groups specified in table below) may be performed as ad-hoc exploratory analyses.

Subgroup	Definition of Group	Analysis Type
Risk	Intermediate 2, High risk	S[(a)]
Sex	Male, Female	S
Race	White, Asian, Other	S
Age	≤65, >65	S

[a] S: Safety (adverse events)

2.1.5. Other General Definitions

2.1.5.1. Individual Study Start/Treatment Start (Day 1), Study Day and Month

Individual Study Start /Treatment start is the date of the Cycle 1, Day 1. If study drug was not administered to a subject, study start is defined as the date of randomization.

The study day is equal to the date of the procedure minus the date of study start plus 1.

A month is defined as 365.25/12 days.

2.1.5.2. End of Treatment, End of Study, Time on Treatment , and Time on Study for Individual Subject

For each individual subject, disposition definitions are defined as below:

End of treatment is defined as the date of the last dose of the study drug. The end of Study is the date of death or the date last known to be alive, defined as the maximum date of the following study evaluations: randomization, vital signs, ECOG performance status, MRIs, physical examination, bone marrow evaluations, labs (hematology, chemistry), study drug administration, subsequent therapy, investigator-assessed PD, relapse from CR, investigator determined response, and survival follow-up.

The treatment cycle is defined as 21 days.

Time on treatment (in weeks) is equal to (the end date of treatment – the date of Cycle 1, Day 1 + 1)/7.

Time on study (in months) is equal to (the end date of follow-up – the date of Cycle 1, Day 1 + 1)/30.4375.

2.1.5.3. Number of Treatment Cycles Received

The number of cycles received is the number of treatment cycles in which there is at least 1 nonzero dose of the study drug.

2.1.5.4. Properties of Treatment Cycles

For each treatment cycle, the following variables are to be derived.

- Cycle delay (≥4 days) — to flag if this cycle is delayed 4 or more days (date of Day 1 in the current cycle – date of Day 1 in the previous cycle is ≥25 days) and checked on the case report form (CRF) by investigator as cycle delayed.
- Reasons for cycle delay (≥4 days) — to record reasons for cycle delay (≥4 days) on the CRF.
- Dose reduction — to flag if the prescribed dose of the study drug is reduced at least once relative to the previous nonzero dose for study drug.
- Reasons for dose reduction — to record reasons for dose reduction on the CRF.
- Dose withholding — to flag if the dose of the study drug is withheld at least once. The number of doses withheld will also be recorded.
- Reasons for dose withholding — to record reasons for dose withholding on the CRF. Reasons for dose withholding will be recorded for each separate dose withholding incident.
- Dose modification — to flag if the dose of the study drug is modified (delayed, reduced, withheld) at least once for study drug.
- Reasons for dose modification - to record reasons for dose modification.
- Number of administered doses — to record the number of nonzero doses actually administered to a subject for study drug.
- Total dose (mg/kg) — to record the sum of nonzero doses actually administered to a subject for study drug.
- Total planned dose (mg/kg) — to record the sum of planned nonzero doses for study drug.
- Dose interval – days between two consecutive dosing dates.

2.1.5.5. Dose Intensity and Relative Dose Intensity

Dose intensity (mg/kg/cycle) will be calculated for study drug by cycle and overall.

The overall dose intensity is equal to the sum of the total dose (mg/kg) received in all cycles divided by the number of cycle received.

The dose intensity per cycle is equal to the sum of the total dose (mg/kg) received in that cycle.

The overall planned dose intensity is equal to the sum of the planned total dose (mg/kg) in all cycles divided by the number of cycles received.

The planned dose intensity per cycle is equal to the sum of the planned total dose (mg/kg) in the cycle.

The relative dose intensity is the ratio of the dose intensity and the planned dose intensity.

2.1.5.6. Age and Age Categories

Age will be calculated at the date of informed consent and integer values will be used. If the date of informed consent is not available, the date of randomization will be used. The following categories of age will be used: ≤65 years, >65 years.

2.1.5.7. Baseline Values and Subject Disease Characteristics

Refer to Section 2.1.1 for the definition of baseline values. Below are some of key disease characteristics.

Bone Marrow Assessments at Baseline

A bone marrow biopsy/aspirate will be performed for all subjects at baseline. Outcomes of bone marrow assessments (Degree of Cellularity, Degree of Fibrosis, and Blasts) will be reported and captured on eCRF.

Myelofibrosis Symptoms (modified MFSAF v2.0)

Myelofibrosis symptoms will be collected at baseline and recorded on the CRF for each item of the questionnaire with possible scores of 0-10.

MF Type

Myelofibrosis type will be collected at baseline and recorded on the CRF as Primary MF, post-ET-MF(PET), and post-PV-MF(PPV).

DIPPS Risk Scores

Dynamic International Prognostic Scoring System (DIPSS) risk score will be collected at baseline and recorded on the CRF (int-2, high).

Time Since Initial Diagnosis

Time since initial diagnosis is defined for all subjects and is calculated from the date of diagnosis to the date of randomization (month).

2.1.5.8. Concomitant Medication

Concomitant medications during study treatment include only those concomitant medication records that are indicated on the CRF as not being prior therapy, and have the onset date, possibly imputed, no earlier than treatment start and no later than treatment end. Concomitant medications include all concomitant medications administered during the study, including those indicated on the CRF at study entry.

2.1.5.9. Protocol Deviations

Subjects with major protocol deviations including eligibility will be listed by treatment arm. Major protocol deviations include: developed withdrawal criteria but not withdrawn, entered but did not satisfy criteria, received a disallowed concomitant treatment, received wrong treatment or incorrect dose, and other.

Protocol deviations will be based on clinical review, but not limited to, the following aspects: (1) eligibility criteria, (2) patient safety, (3) efficacy assessment deviation, (4) treatment compliance. Protocol deviations will be closely monitored during the execution of the study and the final set of protocol deviation criteria will be finalized before database lock.

2.2. Interim Analyses and Data Review Committee

2.2.1. Interim Analyses

There were two interim analyses planned in this study. The first interim review analysis was performed when the first 40 evaluable subjects defined as the subjects who had both baseline and post baseline spleen measurement per investigator MRI, (20 per treatment arm) were followed for at least 12 weeks. The second interim review was performed with a clinical cutoff that was 6 months after the clinical cutoff used for the first interim review.

The goals of the interim reviews were to assess safety and early efficacy and perform comprehensive exposure-response analyses. Based on these analysis, if both doses were efficacious and had an acceptable safety profile, both arms would continue with

enrollment until approximately 100 subjects per arm were enrolled. If only one dose was efficacious and had an acceptable safety profile, then that arm would continue enrollment and enrollment would be stopped in the other arm. In the case that both arms were not efficacious and safe, an alternative dose may be selected based on the PK/PD exposure-response, efficacy, and safety analysis and a protocol amendment would be implemented to reflect this. Study enrollment would continue at the time of interim reviews.

Safety would be evaluated using the following parameters (including but not limited to): adverse events, serious adverse events, dose modifications and subject compliance. PK/PD exposure-response analyses for efficacy parameters (eg, SVR from baseline at Week 12), and for safety parameters (e.g., platelets and hemoglobin) would be conducted. Biomarkers such as change in TA, TL and hTERT from baseline may be evaluated for subjects with available data.

Early efficacy would be evaluated by SVR by MRI and as assessed by the investigator. [redacted] Other efficacy data including symptoms, peripheral blood counts and bone marrow data up to interim will be reviewed as well. This guideline is intended to inform the decision, which is expected to be multifactorial taking into consideration safety, tolerability, and PK/PD modeling.

2.2.2. Data Review Committees

A sponsor Data Review Committee (DRC) was formed to monitor data on a regular basis to ensure the safety of the subjects in this study, assess the evidence of benefit or adverse effects of imetelstat, and to monitor the overall conduct of the study. The DRC would consist of at least one medical expert in the relevant therapeutic area and at least one statistician. At an interim analysis, all available data (e.g. pharmacokinetics [PK], safety, efficacy, pharmacodynamics [PD] biomarkers: TA, TL or hTERT) would be assessed, and integrated exposure-response analyses with PK/PD, safety and efficacy modeling would be conducted. Response data assessed by the investigator up to the cutoff date would be used for the interim review. Based on these analyses, the DRC

would determine if enrollment in one or both treatment arms should continue, or if an alternative dose should be selected for further development. The DRC is responsible for monitoring and reviewing the safety data approximately every 3 months or at an agreed upon review interval until all patients discontinued from study treatment and the post treatment follow up phase, and at the time of interim review stated above. The review consists of at minimum, summaries of overall rates of deaths, serious adverse events (SAEs), and all adverse events and relevant laboratory test data. Details (responsibilities, authorities, procedures) are provided in the DRC Charter.

An independent Hepatic Expert Committee (HEC) was formed to assess/identify hepatic risk in study. The HEC would consist of at least two physicians. The DRC and HEC are not blinded to the treatment throughout the study. A central Independent Review committee (IRC) was formed to assess spleen volume based on MRI scans performed. The IRC would consist of reviewers who are experts in reading MRI scans from Paraxel/Perceptive. The IRC reviewers were blinded to the treatment throughout the study.

2.3. Methods of Analysis

2.3.1. General Analysis Specifications

Unless specified otherwise, summary will be shown by randomized/registered treatment arm with descriptive statistics. The dates used in the derivations are those obtained after applying the rules specified in Section 2.1.2.

The intensity (severity) of adverse events and the grade of abnormal laboratory results will be assessed using National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.0.

2.3.2. Demographics and Baseline Characteristics

All demographic and baseline characteristic variables will be summarized for the ITT population.

The stratification factors used during randomization (palpation, platelet count) will be tabulated.

Demographic and baseline characteristic data will be summarized, including sex, age (years), race, ECOG performance status, region, weight (kg), and height (cm).

Two panels of baseline clinical laboratory tests specified in Section 2.3.9.3 will be summarized: hematology, and serum chemistry. Frequencies of CTC grade at baseline will be reported.

Diagnosis will be summarized by type of Myelofibrosis (PMF, ET-MF[PET], PV-MF[PPV]), risk category (Intermediate-2, High-Risk), relapse criteria (Yes, No for C1-C4), relapse type (5-10 cm with 100% increase, >10 cm with 50% increase, new spleen), age at diagnosis, and time since initial diagnosis (month).

Hepatitis serology test at baseline will be summarized by test type and results.

Myelofibrosis Symptoms at baseline will be summarized using descriptive statistics.

Transfusion history will be summarized by transfusion in last 12 week (yes, no), total transfusion unit, number of transfusion, and transfusion dependency (yes, no)

Medical history will be summarized.

In addition, summaries of bone marrow findings at baseline will be provided.

2.3.3. Prior Systemic and Subsequent Anti-cancer Treatment

Frequency distribution (n and percent) of subjects receiving prior systemic treatment before randomization will be presented, along with medication terms (according to the WHO drug dictionary and ATC). Subsequent anticancer treatment will be summarized in similar fashion as prior systematic treatment.

2.3.4. Prior and Subsequent Radiotherapy

Frequency distribution (n and percent) of subjects receiving radiotherapy before randomization will be provided, along with radiotherapy sites. Subsequent radiotherapy treatment will be summarized in a similar fashion as prior radiotherapy treatment.

2.3.5. Subject Disposition

Disposition information will be summarized for the ITT population.

Subject enrollment will be summarized by country and investigator.

The number of subjects enrolled, treated, still on treatment, and discontinuing study treatment with reasons will be reported for ITT population at the end of treatment.

The number of subjects enrolled, treated, still on study, discontinuing study with reasons will be reported for ITT population at the end of study.

Time on treatment (in weeks) and time on study (in months) will be summarized. Time on study is defined the same way as overall survival (OS) with reversed censoring, i.e., subjects who died will be censored. Based on this definition, time on study is the same as length of follow up. The Kaplan-Meier method will be used to estimate the median time on study.

2.3.6. Extent of Exposure

All exposure summaries are to be presented for the treated analysis set.

The number of treatment cycles received will be summarized along with the duration of treatment exposure, total doses (mg/kg), total number of infusions received, dose interval, dose intensity, and relative dose intensity for study drug. In addition, total doses (mg/kg), dose intensity, and relative dose intensity for study drug will be further summarized by cycle. For study drug, the following summaries are to be reported: number of subjects dosed, number of subjects with dose reduced, dose discontinued, dose withheld, and any change in dosing (reduced, withheld, dose delayed).

Reasons for dose changes will be summarized by cycle, as well as across the study, for each of the following changes: cycle delay (≥4 days), dose reduction, and dose withholding.

For each type of dose changes mentioned above, adverse events leading to changes in doses of study treatment will be summarized.

2.3.7. Concomitant Medications

Concomitant therapies other than antineoplastic agents or other systemic therapies for MF, post ET-MF, and post PV-MF after randomization will be classified by preferred WHO Anatomical Therapeutic Chemical (ATC) and summarized by therapeutic class and pharmacological class. Separate analyses will be performed for concomitant medications started during treatment and for all concomitant medications use (including baseline use).

2.3.8. Efficacy

2.3.8.1. Analysis Specifications

2.3.8.1.1. Level of Significance

All tests will be 1-sided. The primary efficacy endpoint will be tested at the overall significance level of 0.025. No treatment arm comparison will be performed between the two treatment arms.

2.3.8.1.2. Spleen Assessment by Independent Review Committee (IRC)

MRI of the abdomen will be performed to assess spleen volume. The scans will be sent to and read by an Independent Review committee (IRC). The radiological images will be collected at baseline and every 12 weeks up to week 48 and then every 24 weeks and at time of CR/PR.

The radiologists will be blinded to treatment assignment.

2.3.8.1.3. Modified Myelofibrosis Symptom Assessment Form (modified MFSAF v2.0)

Subjects will use a hand-held electronic device to record answers to queries regarding MF symptoms. Questions on these symptoms are from the modified MFSAF v2.0 questionnaire and will be completed by subjects each night beginning at Day -7 (prior to randomization), and continuing to the Week 48 visit. Symptoms will be assessed including 1) night sweats, 2) itchiness, 3) abdominal discomfort, 4) pain under ribs on left side, 5) early satiety, 6) bone or muscle pain, and 7) inactivity. Scoring is from 0 (absent/as good as it can be) to 10 (worst imaginable/as bad as it can be) for each item. The total symptom score (TSS) will be calculated as the 7-day average of daily TSS, which is the summation of 6 individual symptom scores (night sweats, itchiness, abdominal discomfort, pain under ribs on left side, early satiety, and bone or muscle pain).

2.3.8.1.4. Response Assessment Criteria

All assessments will be done according to the modified myeloproliferative neoplasms research and treatment (IWG-MRT) response criteria defined in the protocol (Table 1, Table 2).

Table 1: Response Categories

Response Category	Required Criteria
	For all response categories, benefit must last for ≥12 weeks to qualify as a response
Clinical Improvement (CI)	The achievement of anemia, spleen or symptoms response without progressive disease or increase in severity of anemia, thrombocytopenia, or neutropenia[a]
Anemia Response	Subjects not transfusion dependent at baseline: a ≥ 20 g/dL increase in Hb[b] Transfusion dependent subjects at baseline: becoming transfusion independent. Transfusion independence is defined as absence of any pRBC transfusions for at least 12 "rolling" weeks[b]
Spleen Response	A baseline splenomegaly that is palpable at 5-10 cm, below the LCM, becomes not palpable or A baseline splenomegaly that is palpable at >10 cm, below the LCM, decreases by ≥50% A spleen response requires confirmation by MRI showing ≥35% spleen volume reduction
Symptoms Response	A ≥50% reduction in the modified MFSAF v2.0 TSS[c]
Progressive Disease	Splenomegaly requires MRI showing a ≥25% increase in spleen volume from baseline or Appearance of a new splenomegaly that is palpable at least 5 cm below the LCM or A ≥100% increase in palpable distance, below LCM, for baseline splenomegaly of 5-10 cm or A 50% increase in palpable distance, below LCM, for baseline splenomegaly of >10 cm or 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 X 10(9)/L that lasts for at least 2 weeks
Stable Disease	Belonging to none of the above listed response categories
Relapse	No longer meeting criteria for at least CI after achieving CR, PR, or CI, or Loss of anemia response persisting for at least 1 month or Loss of spleen response persisting for at least 1 month

[a] Increase in severity of anemia constitutes the occurrence of new transfusion dependency or a ≥20 g/L decrease in hemoglobin level from pretreatment baseline that lasts for at least 12 weeks. Increase in severity of thrombocytopenia or neutropenia is defined as a 2-grade decline, from pretreatment baseline, in platelet count or absolute neutrophil count, according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. In addition, assignment to CI requires a minimum platelet count of ≥25,000 X 10^9/L and absolute neutrophil count of ≥0.5 x 10^9/L.

[b] Transfusion dependency before study enrollment is defined as transfusions of at least 6 units of packed red blood cells (PRBC), in the 12 weeks prior to study enrollment, for a hemoglobin level of <85 g/L, in the absence of bleeding or treatment-induced anemia. In addition, the most recent transfusion episode must have occurred in the 28 days prior to study enrollment. Response in transfusion-dependent patients requires absence of any PRBC transfusions during any consecutive "rolling" 12-week interval during the treatment phase, capped by a hemoglobin level of ≥85 g/L.

[c] Symptoms are evaluated by the modified MFSAF v2.0 TSS. The modified MFSAF v2.0 TSS is assessed by the patients themselves and this includes night sweats, itchiness, abdominal discomfort, pain under ribs on left side, early satiety, bone or muscle pain, and inactivity. Scoring is from 0 (absent/as good as it can be) to 10 (worst imaginable/as bad as it can be) for each item. The modified MFSAF v2.0 TSS is the summation of all the individual scores, excluding inactivity (0-60 scale). Symptoms response requires ≥50% reduction in the modified MFSAF v2.0 TSS

For the response categories for CR and PR, the modified 2013 IWG-MRT criteria defined in the protocol will be used (Table 2).

Table 2: Remission Parameters

Parameters	Complete Remission (CR)	Partial Remission (PR) (Scenario 1)	Partial Remission (PR) (Scenario 2)
For all response categories, benefit must last for >12 weeks to qualify as a response			
Bone marrow	Normocellular <5% blasts ≤ Grade 1 fibrosis	Normocellular <5% blasts ≤ Grade 1 fibrosis	Not meeting bone marrow remission criteria
Immature myeloid cells in PB	<2%	<2%	<2%
Hemoglobin	10 g/dL – ULN	8.5 – <10 g/dL – ULN	10 g/dL – ULN
Neutrophils	1 x 10^9/L – ULN	1 x 10^9/L – ULN	1 x 10^9/L – ULN
Platelets	100 x 10^9/L – ULN	50 - <100 x 10^9/L – ULN	100 x 10^9/L – ULN
Spleen	Not palpable and ≤350ml volume	≥35% splenic volumetric reduction by MRI or not palpable	≥35% splenic volumetric reduction by MRI or not palpable
EMH	No non-hepato-splenic EMH	No non-hepato-splenic EMH	No non-hepato-splenic EMH
Symptoms	>70% improvement in symptom score per modified MFSAF v2.0 TSS	>50% improvement in symptom score per modified MFSAF v2.0 TSS	>50% improvement in symptom score per modified MFSAF v2.0 TSS

CR=complete remission; EMH=extramedullary hematopoeisis; MFSAF TSS= Modified Myelofibrosis Symptom Assessment Form Total Symptom Score; MRI=magnetic resonance imaging; PR=partial response; ULN=upper limit of normal

2.3.8.1.5. Spleen and Response Assessment by Investigator

Spleen and response assessment will be performed locally by investigators based on radiological images and using the modified myeloproliferative neoplasms research and treatment (IWG-MRT) response criteria (Table 1, Table 2). In addition, tumor assessment of progressive disease (PD) can also be done based on clinical evidence other than radiological images, which is collected on the CRF.

2.3.8.1.6. Spleen Assessment by IRC

Data collected based on radiological images assessed by the IRC will be evaluated and summarized.

2.3.8.1.7. Best Overall Response

The best overall response is the best response in order of CR/PR/CI/RWCI/SD/PD/NE recorded from the start of the treatment until disease progression or the start of subsequent antineoplastic therapy.

2.3.8.2. Primary Efficacy Endpoints

2.3.8.2.1. Definition and Primary Analysis

The primary endpoints are the spleen response rate based on MRI at Week 24 (spleen volume reduction, SVR at Week 24) and symptom response rate at Week 24 (symptom improvement at Week 24). The SVR at Week 24 is defined as the proportion of subjects who achieve ≥35% reduction in spleen volume at Week 24 and symptom improvement at Week 24 is defined as the proportion of subjects who have ≥50% reduction in total symptom score at Week 24.

The primary analysis of SVR at Week 24 will be based on assessment of MRI by the IRC. The SVR at Week 24 based on assessment of MRI by the investigators will be evaluated and used as supportive evidence.

2.3.8.2.2. Analysis Methods

No treatment comparison between treatment arms for SVR at Week 24 will be performed. The estimates of the SVR at week 24 and symptom improvement at Week 24 in each treatment arm will be presented along with 95% exact confidence intervals based on the treated analysis set.

2.3.8.3. Secondary Efficacy Variables

2.3.8.3.1. Response-Related Endpoints per modified IWG-MRT criteria

Overall Response Rate

Overall response rate (ORR) is defined as the proportion of subjects who achieve complete remission (CR) or partial remission (PR) per modified IWG-MRT criteria based on the overall response assessment for the treated population.

Clinical Improvement

Clinical improvement rate (CI) is defined as the proportion of subjects who achieve CI per modified IWG-MRT criteria based on the overall response assessment for the treated population. Clinical response rate (CRR) is defined as the proportion of subjects who achieve CR, PR, or CI per modified IWG-MRT criteria for the treated population.

Anemia, Spleen, and Symptom Response per Modified IWG-MRT Criteria

Anemia response, spleen response, and symptom response are all defined per modified 2013 IWG-MRT criteria as Table 1.

Analysis Methods

An estimate of the ORR/CRR rate in each treatment arm will be presented along with 95% exact confidence intervals. The number and percentage of subjects falling into each response category will be descriptively tabulated.

CI rate, rate of anemia response, spleen response, and symptom response will be analyzed using the same method as it is done for ORR/CRR.

Best overall response will be tabulated.

The ORR/CRR and other response endpoints such as CI, anemia response, spleen response, symptom response per modified IWG-MRT criteria based on assessment of IRC and investigator will be evaluated and analyzed separately.

2.3.8.3.2. Duration of Response

Definition

Response endpoints include: CR/PR, CR/PR/CI, anemia response, spleen response, and symptom response per modified IWG-MRT criteria. For each of the response endpoints, duration of response will be calculated from the date of initial documentation of a response to the date of first documented evidence of PD or death, whichever occurs first, regardless of the use of subsequent anti-cancer therapy prior to PD or death. Subjects who are progression-free and alive, or have unknown status will be censored at the last disease assessment.

Analysis Methods

The Kaplan-Meier method will be used to descriptively summarize each duration of responses for responders: CR/PR, CR/PR/CI, Spleen, Anemia, and Symptom Response

per modified 2013 IWG-MRT criteria. Analyses based on IRC response and investigator response will be performed separately.

2.3.8.4. Overall Survival

Definition

Overall Survival is measured from the date of Cycle 1, Day 1 to the date of the subject's death. If the subject is alive or the vital status is unknown, OS will be censored at the date that the subject is last known to be alive.

Analysis Methods

The Kaplan-Meier method will be used to estimate the distribution of OS for each treatment arm. One- and two-year overall survival rates, based on Kaplan-Meier estimates, for the 2 treatment arms will be presented, along with their 95% confidence interval.

Analysis based on the ITT population will be the core analysis for OS analysis. In addition, the following analyses for overall survival will be performed:

1) Overall survival time with censoring at dose escalation for subjects who were in Arm B and subsequently dose escalated to 9.4 mg/kg

2) Overall survival time with Rank Preserving Structural Failure Time (RPSFT) Method and Overall survival time with Inverse Probability of Censoring Weighting (IPCW) Method

3) Overall survival time excluding subjects who were in Arm B and subsequently dose escalated to 9.4 mg/kg

4) Overall survival time excluding subjects who were enrolled on or after Sept 9, 2016, when at the time study enrollment was suspended. Those subjects who were already in screening phase and subsequently determined to be eligible for the study, were allowed to proceed with treatment assignment to the 9.4 mg/kg arm per investigator discretion, subject agreement, and IRB/IEC notification.

2.3.8.5. Exploratory Analyses

To explore the potential benefit, efficacy signal of imetelstat on MF patients, exploratory analyses may be conducted include but not limited to:

- TSS score, spleen volume and spleen length will be summarized over time, along with plots.
- Waterfall plots of TSS score, spleen volume, and spleen length collected at Week 24 will be generated to evaluate pattern of changes from baseline.
- Time to spleen response, which is defined as time from date of the Cycle 1, Day 1 to the date of the first documented spleen response (SVR >=35%), will be descriptively summarized for subjects who achieve spleen response.
- Time to symptom response, which is defined as time from date of the cycle 1, Day 1 to the date of the first documented symptom response (TSS>=50%), will be descriptively summarized for subjects who achieve symptom response.
- Time to anemia response will be descriptively summarized for anemia responders per modified IWG-MRT criteria.
- Improvement of fibrosis grade is defined as a fibrosis grade better than baseline fibrosis grade. The proportion of subjects who achieve improvement of fibrosis grade for the treated population will be summarized.

For interim analysis, Week 12 data will be used instead of Week 24 when it is applicable, including, but not limited to, leukemia free survival, bone marrow (fibrosis grade, blasts, cellularity), transfusion reduction/ transfusion independence.

To explore the potential benefit related to transfusion, the analyses of rate of 12-week RBC TI response and time to 12-week RBC TI response will be performed.

2.3.8.6. Patient-Reported Outcomes (PRO)

2.3.8.6.1. EORTC QLQ-C30

EORTC QLQ-C30 includes 30 separate items resulting in 6 functional scales (global health status, physical functioning, role functioning, emotional functioning, cognitive functioning, and social functioning) plus 3 symptom scales (fatigue, nausea and vomiting, and pain), Scores will be derived using validated scoring algorithms according to EORTC QLQ-C30 Scoring Manual (3rd Edition, 2001).

2.3.8.6.2. EQ-5D-5L

The EQ-5D is a 5-item questionnaire and a “thermometer” visual analogue scale ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). The scores for the 5 separate questions are categorical and should not be analyzed as ordinal numbers.

The scores for the 5 categorical dimensions will be used to compute a single utility score ranging from zero (0.0) to 1 (1.0) representing the general health status of the individual. The United Kingdom weights will be used to generate patient utilities from the 5 dimensions of the EQ-5D in this study.

2.3.8.6.3. Brief Pain Inventory (BPI)

The BPI is a 7-item questionnaire.

2.3.8.6.4. Patient's Global Impression of Change (PGIC)

The PGIC is a questionnaire form with scores ranging from 1 (Very much improved) to 7 (Very much worse).

2.3.8.6.5. Analysis Methods for PRO

Descriptive statistics (N, mean, SD, median, minimum, maximum) at baseline and each post-baseline time point is reported by treatment arm for each scale of EORTC QLQ-C30, EQ 5D visual analogue scale and weighted utility score, BPI, and PGIC. Analyses of compliance will be performed for each PRO measures.

The PRO measure of interest is EORTC-QL-C30 and BPI. Value of threshold for improvement/worsening is estimated by distribution based method. Specifically, improvement/ worsening is defined as ≥ half of standard deviation at baseline points for each assessment post baseline (add ref Norman 2003). Rate of improvement/worsening, and time to improvement/worsening for EORTC-QL-C30 and BPI will be summarized for each treatment arm.

2.3.8.7. Biomarker

Biomarker parameters of Telomerase Activity (TA), Telomere Length (TL), hTERT RNA Expression (hTERT), molecular response and Cytogenetic response will be assessed at day 1, day 8 of Cycle 1, day 1 of cycle 2, at time of CR/PR, and at time of PD. The change and percent change of each parameter (TA, TL, and hTERT) will be summarized (N, mean, SD, minimum, maximum) at each scheduled assessment visit, along with plots. Cytogenetic response rate and molecular response will be derived according to IWG-MRT criteria and will be summarized based on biomarker evaluable population. A separate biomarker SAP may be prepared, and results of biomarker analyses may be presented in a separate report.

2.3.9. Safety

The safety of the treatment arms will be summarized using laboratory test results, vital sign measurements, physical examination results, and the incidence and severity of adverse events.

2.3.9.1. Adverse Events

Treatment-emergent adverse events are defined as any adverse event that occurs after administration of the first dose of study drug and through 30 days after the last dose of study drug, any event that is considered drug-related (possibly, probably, or possibly related) regardless of the start date of the event, or any event that is present at baseline but worsens in intensity or is subsequently considered drug-related by the investigator. The severity of adverse events will be assessed using NCI CTCAE Version 4.0.

The Medical Dictionary for Regulatory Activities (MedDRA version 20.0) dictionary will be used to code the investigator's adverse event terms to preferred term and system organ class. Adverse event tables will summarize incidence by the treatment arm actually received.

For each AE, its relationship to study drug is determined by the investigator and recorded on the CRF. An adverse event is considered as related to study drug if it is possibly, probably, or very likely related to study drug imetelstat.

An overview of treatment-emergent adverse events as well as the incidence of treatment-emergent adverse event will be summarized. The following tabulations will also be presented:

- Incidence of treatment-emergent adverse events by system organ class and preferred term.
- Incidence of treatment-emergent adverse events, especially the most commonly reported events, ie, those events reported by at least 5% of subjects, by MedDRA system organ class and preferred term.
- Incidence of treatment-emergent adverse events considered by the investigator to be related to study drug, by system organ class and preferred term.
- Incidence of treatment-emergent adverse events by CTCAE toxicity grade.
- Incidence of treatment-emergent adverse events leading to withdrawal from treatment by system organ class and preferred term (adverse events leading to discontinuation of all study drugs).
- Incidence of treatment-emergent adverse events leading to death by system organ class and preferred term.

For treatment-emergent adverse events with CTCAE toxicity Grade 3, 4, or 5, the following summaries are to be produced:

- Incidence of Grade 3, 4, or 5 treatment-emergent adverse events, by system organ class and preferred term.
- Incidence of Grade 3, 4, or 5 treatment-emergent adverse events considered by the investigator to be related to study drug, by system organ class and preferred term.

For treatment-emergent serious adverse events, the following summaries are to be produced:

- Incidence of treatment-emergent serious adverse events, by system organ class and preferred term.
- Incidence of treatment-emergent serious adverse events considered by the investigator to be related to study drug, by system organ class and preferred term.
- Incidence of treatment-emergent serious adverse events by relationship to study drug.

The subgroup analyses of adverse events will be done as specified in Section 2.1.3.5.

Tabulation for adverse events of interest (hepatic AE, site reaction AE) will be provided. Details will be specified in Data Presentation Specifications (DPS).

2.3.9.2. Death

Tabulations will be produced for all-cause mortality 30 and 60 days after the first dose, and within 30 days after the last dose, as well as all deaths for all subjects treated. The cause of death will also be summarized in this table. In particular, frequencies of deaths that are due to study treatment-related adverse events will also be reported.

2.3.9.3. Clinical Laboratory Tests

Laboratory data of hematology and clinical chemistry up to 30 days after last dose will be reported in SI units.

Summary statistics (N, mean, SD, median, and range) will be calculated for the raw data and for their changes from baseline by scheduled visits, as well as for last value and for the changes from baseline to the last value. Within a cycle, summary statistics from the worst values will be provided. Individual values outside the normal ranges will be identified (by “H” for high and “L” for low) in the data listings displaying the absolute values for each subject.

Displays of over-time summaries will be presented for the following key laboratory parameters: hemoglobin, white blood cell count, absolute neutrophil count, platelet count, manual percent immature myeloid cells (blasts, promyelocytes,

myelocytes,metamyelocytes and nucleated RBC), sodium, alkaline phosphatase, potassium, lactic acid dehydrogenase, blood urea nitrogen, albumin, creatinine, serum ferritin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin (with fractionation if abnormal), INR(or PT) and aPTT • INR (or PT) and aPTT.

Shift tables for each cycle will be produced for selected laboratory parameters, to include hemoglobin, white blood cell count, neutrophils, platelet count, creatinine, total bilirubin, ALT, AST, lactic dehydrogenase, sodium, and potassium. These tables will summarize by cycle the number of subjects with each baseline CTC grade and changes to the maximum CTCAE grade in the cycle. For laboratory parameters without CTCAE grade, shifts from baseline to highest per cycle (Low, Normal, High) categories will be used.

Shift tables from baseline to worst value on treatment (from treatment start to 30 days after last dose or the end of treatment visit date, whichever is later) will also be provided. The worst toxicity grade during the treatment will be tabulated.

In addition, incidence of persistent (≥2 or 4 weeks) and severe (≥3 or 4 grade) cytopenia (thrombocytopenia/neutropenia) will be evaluated. Maximum post-baseline CTCAE grade for anemia, leukopenia, neutropenia, and thrombocytopenia will be tabulated.

2.3.9.4. ECOG Performance Status

The ECOG Performance Status scores will be summarized over time. Change in ECOG scores from baseline will also be summarized.

REFERENCES

1. Berger, RL. Multiparameter hypothesis testing and acceptance sampling. *Technometrics*. 1982;24:295-300
2. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549-556
3. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993;11:570-9..
4. Yellen SB, Cella DF, Webster K, et al. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J. Pain Symptom Manage 1997;13:63-74
5. Norman GR, Sloan JA, Wyrwich KW (May 2003). "Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation". Med Care 41 (5): 582–92.