



Statistical Analysis Plan

NCT Number: NCT05210790

Title: A Phase 3 Study of the Hepcidin Mimetic Rusfertide (PTG-300) in Patients with Polycythemia Vera

Study Number: PTG-300-11

Document Version and Date: Version 1.0, 18 Feb 2025

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

For non-commercial use only

9. DOCUMENTATION OF STATISTICAL METHODS

- [Statistical Analysis Plan \(SAP\) for PTG-300-11, Version 1.0, 18 February 2025.](#)

For non-commercial use only

STATISTICAL ANALYSIS PLAN (SAP)

Study Title	A Phase 3 Study of the Hepcidin Mimetic Rusfertide (PTG-300) in Patients with Polycythemia Vera
Test Drug	Rusfertide (PTG-300)
Study Protocol Number	PTG-300-11
IND Number:	152414
NCT Number:	NCT05210790
EU Clinical Trial Number:	2023-509750-58-00
Document Version:	Protocol Amendment 5.1 dated 09 January 2025 Protocol Amendment 5 dated 19 November 2024 Protocol Amendment 4 dated 06 September 2024 Protocol Amendment 3.1 dated 25 July 2023 Protocol Amendment 3 dated 30 June 2023 Protocol Amendment 2 dated 01 November 2022 Protocol Amendment 1 dated 03 December 2021 Original Protocol, 05 November 2021
Sponsor:	Protagonist Therapeutics, Inc. 7707 Gateway Boulevard, Suite 140 Newark, CA 94560-1160

Version	Date
1.0	18 February 2025

TABLE OF CONTENTS

1.	SIGNATURE PAGE	6
2.	LIST OF ABBREVIATIONS.....	8
3.	INTRODUCTION	10
4.	STUDY OBJECTIVES AND ENDPOINTS.....	11
4.1.	Objectives	11
4.2.	Efficacy Endpoints.....	11
4.3.	Safety Endpoints	14
4.4.	Pharmacokinetic (PK)/Pharmacodynamic (PD) Endpoints.....	14
5.	INVESTIGATIONAL PLAN.....	15
5.1.	Overall Study Design and Plan.....	15
5.2.	Investigational Treatment	15
5.3.	Treatment Duration Including Post-Treatment Safety Follow-up.....	16
5.4.	Treatment Blinding and Unblinding.....	16
5.5.	Randomization	16
5.6.	Subject Withdrawal Criteria	16
6.	DETERMINATION OF SAMPLE SIZE AS IN THE PROTOCOL.....	18
6.1.	Primary Endpoint.....	18
6.2.	Key Secondary Endpoints.....	18
6.2.1.	Key Secondary Endpoint #1: Mean Number of Phlebotomies Between Baseline Through Week 32 (Inclusive)	18
6.2.2.	Key Secondary Endpoint #2: Proportion of Subjects with All Hematocrit Values <45% Between Baseline Through Week 32 (Inclusive)	18
6.2.3.	Key Secondary Endpoint #3: Mean Change from Baseline in Total Fatigue Score Based on PROMIS Short Form 8a at Week 32	19
6.2.4.	Key Secondary Endpoint #4: Mean Change from Baseline in Total Score Based on MFSAF v4.0 at Week 32	19
7.	GENERAL ANALYSIS CONSIDERATIONS	20
7.1.	General Rules.....	20
7.2.	Specific Rules for Questionnaires	20
8.	NOTATION OF DOSE LEVEL FOR STUDY DRUG AND VISITS.....	23
8.1.	Notation of Dose Levels for Study Drug.....	23
8.2.	Analysis Epochs.....	24
9.	BASELINE, STUDY DAY, AND TIME WINDOWS.....	26

9.1.	Baseline for Study.....	26
9.1.1.	PK, PD, and ECG Parameters.....	26
9.1.2.	MFSAF	26
9.2.	Part 1b: Week 32	26
9.2.1.	MFSAF	26
9.3.	Part 2: Week 52	26
9.3.1.	MFSAF	26
9.4.	Change from Baseline.....	26
9.5.	Study Day	27
9.6.	Analysis Visits.....	27
10.	ANALYSIS POPULATIONS	28
10.1.	Intent-to-Treat Set (ITT).....	28
10.2.	Full Analysis Set (FAS).....	28
10.3.	Safety Analysis Set (SAS).....	28
10.4.	Per Protocol Set (PPS).....	28
10.5.	PK Analysis Set (PKAS)	28
10.6.	PD Analysis Set (PDAS)	28
11.	DESCRIPTION OF STUDY POPULATION.....	29
11.1.	Subject Disposition.....	29
11.2.	Enrollment	29
11.3.	Demographics and Baseline Characteristics.....	29
11.4.	Medical History	30
11.5.	Physical Examination	30
11.6.	Protocol Deviations	31
12.	EFFICACY ANALYSES	32
12.1.	General Consideration	32
12.2.	Analysis of Efficacy Endpoints	33
12.2.1.	Primary Efficacy Endpoint	33
12.2.2.	Key Secondary Efficacy Endpoints.....	34
12.2.3.	Analysis of Other Secondary Endpoints.....	35
12.3.	Handling of Dropouts, Missing Data, Intercurrent Events, and Non-Responder Categories	40
12.4.	Sensitivity Analysis of the Primary and Key Secondary Endpoints.....	44

12.4.1.	Per Protocol Analysis	44
12.4.2.	Full Analysis Set Analysis	44
12.4.3.	Tipping Point Analysis	44
12.4.4.	Assessing Normality Assumptions of ANCOVA and MMRM Model	46
12.5.	Multiplicity Adjustment	47
12.6.	Examination of Subgroups	48
13.	CLINICAL PHARMACOLOGY ANALYSIS	49
13.1.	Pharmacokinetic (PK) Analysis	49
13.2.	Pharmacodynamic (PD) Analysis	49
13.3.	Exploratory Biomarkers of Interest	49
14.	SAFETY ASSESSMENTS	50
14.1.	Extent of Exposure and Compliance	50
14.2.	Adverse Events	51
14.2.1.	Adverse Event Dictionary	51
14.2.2.	Adverse Event Severity	51
14.2.3.	Relationship of Adverse Event to Study Drug	51
14.2.4.	Serious Adverse Event	51
14.2.5.	Treatment-Emergent Adverse Events	51
14.2.6.	Summary of Adverse Events	52
14.3.	Clinical Laboratory Tests	54
14.3.1.	Hematology	55
14.3.2.	Serum Chemistry	56
14.3.3.	Urinalysis	56
14.3.4.	Additional Lab Tests	57
14.4.	Prior and Concomitant Medications	57
14.5.	Vital Signs	57
14.6.	12-Lead ECG	58
14.6.1.	Change in Spleen Size from Screening	59
14.7.	Immunogenicity	59
15.	REFERENCES	60
APPENDIX 1. MISSING OR INCOMPLETE DATA IMPUTATION		61
APPENDIX 2. SAMPLE SAS CODES FOR MULTIPLE IMPUTATION		62

APPENDIX 3. DEFINING A SYMPTOMATIC SUBGROUP OF PATIENTS FOR
THE PV-FOCUSED PRO SECONDARY ENDPOINTS 67

LIST OF TABLES

Table 1:	EORTC Scales	21
Table 2:	Notation of Dose Levels for Study Drug.....	23

LIST OF FIGURES

Figure 1:	Study Design PTG-300-11	15
Figure 2:	Severity Cut-off for the PROMIS Fatigue Short Form 8a Total T-score in the “Moderate to Severe” Subgroup of Patients Using the Descriptive Analysis of the Baseline PROMIS Fatigue Short Form 8a Total T-score.....	69
Figure 3:	Severity Cut-off for the MFSAF TSS4 in the “Moderate to Severe” Subgroup of Patients Using the Descriptive Analysis of the MFSAF TSS4 Baseline Score	70
Figure 4:	Severity Cut-off for the MPN-SAF Problems with Concentration Symptom Score in the “Moderate to Severe” Subgroup of Patients Using the Descriptive Analysis of the Baseline MPN-SAF Problems with Concentration Symptom Score.....	71

1. SIGNATURE PAGE

Authored by:

[REDACTED]

18-Feb-2025 | 13:40 PST

[REDACTED], PhD
[REDACTED], Biometrics

Date

Reviewed and approved by:

[REDACTED]

18-Feb-2025 | 14:20 PST

[REDACTED], PhD
[REDACTED], Biometrics

Date

[REDACTED]

18-Feb-2025 | 15:13 PST

[REDACTED], MS
[REDACTED], Statistical Programming

Date

[REDACTED]

18-Feb-2025 | 14:03 PST

[REDACTED], MD
[REDACTED], Pharmacovigilance

Date

[REDACTED]

18-Feb-2025 | 14:11 PST

[REDACTED], PhD
[REDACTED], Clinical Development

Date

[REDACTED]

18-Feb-2025 | 14:53 PST

[REDACTED], MD, MS

Date

For non-commercial use only

2. LIST OF ABBREVIATIONS

Abbreviation	Full Notation
ADA	Anti-drug antibodies
AE	Adverse event
AESI	Adverse events of special interest
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ANCOVA	Analysis of Covariance
AR(1)	Autoregressive of order 1
ARH(1)	Heterogeneous autoregressive of order 1
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
BLQ	Below the limit of quantification
BMI	Body mass index
°C	Degree Celsius
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CRO	Contract research organization
CS	Clinically significant
CSR	Clinical study report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DBP	Diastolic blood pressure
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EDC	Electronic Data Capture
eGFR	Estimated glomerular filtration rate
EMA	European Medicines Agency
EORTC-QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire
EOS	End of Study
EOT	End of Treatment
ET	Early Termination
FAS	Full analysis set
FMQ	FDA Medical Queries
HBsAG	Hepatitis B Surface Antigen
HBc	Hepatitis B core
HCT	Hematocrit
HGB	Hemoglobin
HIV	Human Immunodeficiency virus
HLGT	High-level group term
HLT	High-level term
ICH	International Council for Harmonisation
INR	International normalized ratio
ISR	Injection site reaction
IRT	Item-response theory
ITT	Intent to treat
IxRS	Interactive response system
JAK2	Janus kinase 2

Kg	Kilogram
LDH	Lactate dehydrogenase
LLT	Lowest-level term
LSMEAN	Least squares mean
Max	Maximum
MAR	Missing at random
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
MCV	Mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
MFSAF	Myelofibrosis Symptom Assessment Form version 4.0
Min	Minimum
MMRM	Mixed model for repeated measures
MNAR	Missing not at random
MPN-SAF	Myeloproliferative Neoplasm Symptom Assessment Form
MRI	Magnetic resonance imaging
PAP	Psychometric Analysis Plan
PD	Pharmacodynamics
PDAS	PD analysis set
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PK	Pharmacokinetics
PKAS	PK analysis set
PPS	Per protocol set
PROMIS	Patient Reported Outcomes Measurement Information System
PT	Preferred term, prothrombin time
PTT	Partial thromboplastin time
PV	Polycythemia vera
QTc	Corrected QT interval
QTcF	Corrected QT interval according to Fridericia's formula
RBC	Red blood cell
RDW	Red cell distribution width
ReML	Restricted maximum likelihood
ROC	Receiver Operating Characteristics
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Safety analysis set; Statistical Analysis System
SBP	Systolic blood pressure
SD	Standard deviation
SI	International system of unit
SOC	System organ class
TEAE	Treatment-emergent adverse event
TLFs	Tables, listings, and figures
TOEPH	Heterogeneous Toeplitz
TSAT	Transferrin saturation
TSS4	Total Symptom Score – 4 items. The sum of four items of the MFSAF v4.0 (fatigue, itching, night sweats, and abdominal discomfort)
TSS7	Total Symptom Score – 7 items. The sum of seven items of the MFSAF v4.0
WBC	White blood cell
WHO	World Health Organization

3. INTRODUCTION

The purpose of the Statistical Analysis Plan (SAP) is to document the statistical methods to be implemented during the analyses of data collected within the scope of Protagonist Therapeutics, Inc. protocol PTG-300-11, A Phase 3 Study of the Heparin Mimetic Rusfertide (PTG-300) in Patients with Polycythemia Vera. The plan will provide specific guidelines from which the statistical analyses will proceed.

This SAP is based on the study protocol amendment 5.1 dated 09 January 2025, the associated electronic case report forms (eCRFs), and ClinTrak EDC eDiary. The SAP may be updated, as necessary, in case of protocol amendments, updates to data collected in the study and/or data handling rules. The SAP will be finalized before unblinding. For any differences in the planned analyses between the SAP and the protocol, the SAP will take precedence. Any deviations from this plan will be documented in either an SAP amendment or in the clinical study report (CSR).

PTG-300 and rusfertide are used interchangeably in this document.

For non-commercial use only

4. STUDY OBJECTIVES AND ENDPOINTS

4.1. Objectives

To evaluate the safety and efficacy of rusfertide treatment in subjects with polycythemia vera in maintaining hematocrit control.

4.2. Efficacy Endpoints

Primary endpoint:

The proportion of subjects achieving a response starting at Week 20 through Week 32 (inclusive) who receive rusfertide compared to placebo. Response is defined as absence of phlebotomy eligibility.

Phlebotomy eligibility is defined as **either**:

- A confirmed hematocrit $\geq 45\%$ and that is at least 3% higher than the baseline hematocrit. Confirmation is defined as 2 consecutive hematocrit assessments that are $\geq 45\%$ and at least 3% higher than the baseline hematocrit.
- or**
- a hematocrit $\geq 48\%$.

Hematocrit values from each clinical site's local laboratory will be used for the primary endpoint.

Key secondary endpoints: Comparison of rusfertide to placebo

1. Mean number of phlebotomies between Baseline through Week 32 (inclusive).
2. Proportion of subjects with all hematocrit values $< 45\%$ between Baseline through Week 32 (inclusive). A single hematocrit value $\geq 45\%$ is allowed. Hematocrit values from each clinical site's local laboratory will be used for this endpoint.
3. Mean change from baseline at end of Part 1a (Week 32) in the Patient Reported Outcomes Measurement Information System (PROMIS) Fatigue Short Form 8a total T-score.
4. Mean change from baseline at end of Part 1a (Week 32) in the Myelofibrosis Symptom Assessment Form version 4.0 (MFSAF) Total Symptom Score-7 items (TSS7). TSS7 is the sum of 7 items of the MFSAF. The weekly TSS7 at Week 32 will be calculated by averaging daily TSS7 over a 7-day interval leading to Week 32 (inclusive).

Other secondary endpoints:

Part 1a - Baseline through Week 32 (inclusive):

Comparison of rusfertide to placebo during randomized, double-blind, placebo-controlled, add-on parallel phase (Baseline to Week 32)

PV-focused PRO-related secondary endpoints in subjects with moderate or severe symptoms at baseline:

- Mean change from baseline at end of Part 1a (Week 32) in the PROMIS Fatigue Short Form 8a total T-score among patients who have moderate or severe symptoms at baseline.
- Mean change from baseline at end of Part 1a (Week 32) in the MFSAF Total Symptom Score-4 items (TSS4) among patients who have moderate or severe symptoms at baseline. TSS4 is the sum of 4 items of the MFSAF (fatigue, itching, night sweats, and abdominal discomfort). The weekly TSS4 at Week 32 will be calculated by averaging daily TSS4 over a 7-day interval leading to Week 32 (inclusive).
- Mean change from baseline at end of Part 1a (Week 32) in the Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF) Problems with Concentration among patients who have moderate or severe symptoms at baseline.
- Proportion of subjects with a meaningful within-person change at Week 32 in the PROMIS Fatigue SF8a T-score, MFSAF TSS4, and MPN-SAF problems with concentration among subjects who have moderate or severe symptoms at baseline.

Other PRO-related secondary endpoints:

- Mean change from baseline at each scheduled assessment in individual questions and the total fatigue T-score based on PROMIS® Fatigue Short Form 8a in all Intent-to-Treat (ITT) subjects.
- Mean change from baseline at each scheduled assessment in individual questions and the total fatigue T-score based on PROMIS® Fatigue Short Form 8a for subjects who have moderate or severe symptoms at baseline.
- Mean change from baseline at each scheduled assessment in individual symptom score and the MFSAF v4.0 TSS4 for all ITT subjects.
- Mean change from baseline at each scheduled assessment in individual symptom score and the MFSAF v4.0 TSS4 for subjects who have moderate or severe symptoms at baseline.
- Proportion of subjects with a meaningful within-person change at Week 32 in the PROMIS Fatigue SF8a T-score, MFSAF TSS4, and MPN-SAF problems with concentration among all ITT subjects.
- Mean change from baseline at each scheduled assessment in individual symptom score and the total score based on MPN-SAF TSS 10.

- Mean change from baseline in Patient Global Impression of Severity (PGI-S) at each scheduled assessment.
- Proportion of subjects with at least one unit improvement in the PGI-S from baseline, i.e., a ≥ 1 point decrease in the PGI-S, e.g., a change from “Moderate (2)” to “Mild (1)” or “Mild (1)” to “None (0)” at each scheduled visit. Similarly, proportion of subjects with at least one unit worsening in the PGI-S from baseline.
- Proportion of subjects with a rating on the Patient Global Impression of Change (PGI-C) of either “much better” or “a little better” at each scheduled assessment.
- Mean PGI-C at each scheduled assessment.
- Mean change from baseline at each scheduled assessment in individual symptom score and the MFSAF v4.0 Total Symptom Score-7 items (TSS7).
- Mean change from baseline in the individual domain scores and the total European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) score at Week 32.

Hematocrit and Phlebotomy-related secondary endpoints:

- Median time to first hematocrit $\geq 45\%$ after randomization.
- Median time to phlebotomy-eligible after randomization.
- Proportion of subjects with at least one hematocrit $\geq 48\%$.
- Proportion of subjects who maintain absence of phlebotomy eligibility.
- Phlebotomy annualized rate.

Part 1a+1b: Baseline through Week 52

- Proportion of subjects originally randomized to rusfertide achieving absence of phlebotomy eligibility (durable response) for 52 weeks.
- Median time to first hematocrit $\geq 45\%$ during Part 1b excluding Part 1a.
- Median time to first phlebotomy during Part 1b excluding Part 1a.
- Proportion of subjects with all hematocrit values $< 45\%$ between Baseline through Week 52 (inclusive). Two hematocrit values $\geq 45\%$ are allowed. Hematocrit values from each clinical site’s local laboratory will be used for this endpoint.
- Proportion of subjects randomized to placebo who meet the phlebotomy eligibility in Part 1a and Part 1b, separately.

Part 1a+1b+2 – Baseline through Week 156

- Endpoints from Baseline through Week 32 and Baseline through Week 52 above will be repeated to assess the long-term efficacy, as appropriate.

4.3. Safety Endpoints

Part 1a+1b+2: Baseline through Week 156

- Overall summary for treatment-emergent adverse events (TEAEs), serious TEAEs and AEs of special interest (AESI).
- Clinical safety laboratory tests, physical examination, vital signs (heart rate, systolic and diastolic blood pressure, respiratory rate) and 12-lead electrocardiogram (ECG) findings will be summarized.

4.4. Pharmacokinetic (PK)/Pharmacodynamic (PD) Endpoints

PK endpoints are predose and postdose rusfertide concentrations.

PD endpoints are predose and postdose serum iron, serum ferritin, transferrin saturation (TSAT), and mean corpuscular volume (MCV).

For non-commercial use only

5. INVESTIGATIONAL PLAN

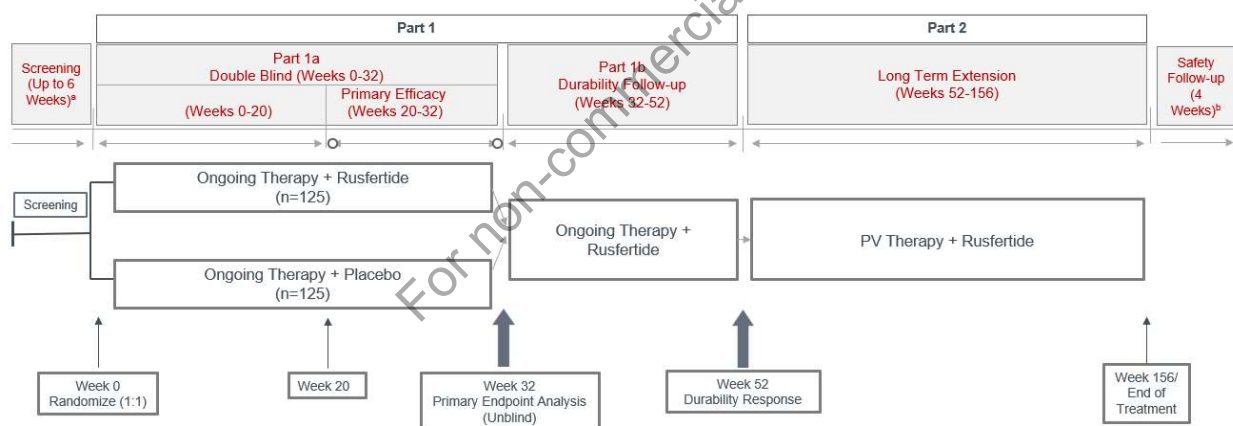
5.1. Overall Study Design and Plan

This is a Phase 3, multicenter, global randomized trial that compares the efficacy and safety of rusfertide compared to placebo when added on to current therapy for polycythemia vera (PV) (Figure 1). The study population is subjects with PV who require frequent phlebotomies to control their hematocrit. Approximately 250 subjects will be randomized.

This Phase 3 study in subjects with polycythemia vera is designed as follows:

- Part 1a: randomized, double-blind, placebo-controlled, add-on parallel-group period for 32 weeks. Subjects will be randomized 1:1 in a blinded fashion to 32 weeks of add-on rusfertide or placebo treatment. Rusfertide or placebo will be added-on to each subject's ongoing therapy for PV. Randomization will be stratified by ongoing PV therapy.
- Part 1b: open-label treatment phase during which all subjects who complete Part 1a successfully will receive rusfertide for 20 weeks (Week 32 through Week 52).
- Part 2: long term extension for 104 weeks during which all subjects who complete Part 1b will continue to receive rusfertide.

Figure 1: Study Design PTG-300-11



^a Phlebotomy history for up to 52 weeks.

^b In addition, cancer event follow-up will be conducted at approximately 6 months and 12 months after the last dose of rusfertide to determine if there has been a development of new cancer(s).

5.2. Investigational Treatment

In Part 1a, subjects are randomized (1:1) to:

- Treatment Arm 1: Rusfertide added to ongoing PV therapy at randomization.
- Treatment Arm 2: Placebo added to ongoing PV therapy at randomization.

In Part 1b, all subjects receive open-label rusfertide added to the ongoing PV therapy at randomization.

In Part 2, all subjects receive open-label rusfertide. The ongoing PV therapy may be changed at the investigator's discretion.

5.3. Treatment Duration Including Post-Treatment Safety Follow-up

Individual subjects will participate in this study for up to 166 weeks as follows:

- Screening: Up to 6 weeks
- Part 1a: Double-blind: 32 weeks (Baseline to Week 32)
- Part 1b: Open-label: 20 weeks (Week 32 to Week 52)
- Part 2: Long-term extension: 104 weeks (Week 52 to Week 156)
- Post-treatment safety follow-up: 4 weeks

Subjects will be contacted at approximately 6 months and 12 months after the last dose of rusfertide to determine if there has been a development of new cancer(s).

5.4. Treatment Blinding and Unblinding

During Part 1a of the study, the Sponsor, contract research organizations (CROs), Medical Monitors, Investigator, site pharmacist (or designee), site personnel involved in study conduct, and subjects will remain blinded to study medication assignment. Treatment assignment will be implemented using an interactive response system (IxRS).

Treatment assignment for an individual subject in Part 1a may be made available to the Investigator (or designee) only in the event of a medical emergency or an AE that necessitates identification of the clinical trial material for the welfare of that subject (see protocol Section 13.6). If an investigator unblinds a subject's treatment assignment before the subject completes Part 1a of the study, the subject will be discontinued from the study.

Subjects who complete Part 1a will continue to Part 1b. Since a subject randomized to placebo in Part 1a will receive 20 mg of active drug in Part 1b, the site and subject will be unblinded. Because of potential unblinding at the start of Part 1b, all required assessments at Week 32 must be completed prior to continuing to Part 1b.

Treatment in Parts 1b and 2 is not blinded.

5.5. Randomization

Subjects will be stratified by ongoing PV therapy (phlebotomy only, phlebotomy+hydroxyurea, phlebotomy+ruxolitinib, phlebotomy+interferon, phlebotomy+other) and randomized 1:1 to receive rusfertide or placebo for 32 weeks using the IxRS. Study drug (rusfertide or placebo) will be added on to each subjects' ongoing PV treatment at randomization.

5.6. Subject Withdrawal Criteria

A subject may choose to withdraw at any time. A subject should be withdrawn if the Investigator or Sponsor thinks withdrawal is in the subject's best interest.

A subject must be withdrawn if:

- Subject requests withdrawal from the study.
- Subject is noncompliant with study drug administration or study procedures.
- Subject becomes pregnant.
- Clinically significant adverse events which prevent continuation of the study medication, as determined by the Investigator.
- Subject starts a new cytoreductive therapy in either Part 1a or Part 1b.
- Subject receives cytoreductive therapy other than hydroxyurea, interferon, ruxolitinib or an approved JAK inhibitor.
- Clinically significant laboratory values which require change in PV therapy, as determined by the Investigator.
- Any Common Terminology Criteria for Adverse Events (CTCAE) Grade 3 or Grade 4 toxicity, when the Investigator, in consultation with the medical monitor, considers that it is in the subject's best interest to discontinue treatment with study drug.
- Subject starts an experimental therapy.
- Subject's treatment is unblinded before the subject completes Part 1a.
- Lost to follow-up (a subject will be considered lost to follow-up if he/she repeatedly fails to return for scheduled visits).

6. DETERMINATION OF SAMPLE SIZE AS IN THE PROTOCOL

Overall, the study with approximately 250 subjects, randomized 1:1 to treatment arms (125 per arm) will have at least 95% power to test the primary endpoint. The primary and key secondary endpoints will be tested using $\alpha=0.05$ (2-sided), with overall control for Type 1 error rate.

6.1. Primary Endpoint

The response rate estimate in the RESPONSE trial (Vannucchi 2015) was 19.6% for Best Available Therapy and 60% for ruxolitinib. In order not to overestimate the treatment effect for this trial, a response rate (absence of phlebotomy-eligibility) of approximately 25% on placebo and 50% from rusfertide during Weeks 20 to 32 (inclusive) was assumed. Based on these assumptions 95 subjects will be required per treatment group to detect a difference in the proportion of subjects treated with rusfertide achieving clinical response compared to subjects treated with placebo at a 0.05 level of significance (2-sided) with approximately 95% power using a z-test based on the large sample normal approximation to binomial distribution.

This endpoint will be presented as the first key secondary efficacy endpoint for European Medicines Agency (EMA) submissions.

6.2. Key Secondary Endpoints

6.2.1. Key Secondary Endpoint #1: Mean Number of Phlebotomies Between Baseline Through Week 32 (Inclusive)

The objective is to evaluate the efficacy of rusfertide in maintaining hematocrit control. Based on available data from a Phase 2 study, PTG-300-04, it is observed during the open-label dose-finding part of the study, the mean number of phlebotomies in 16 weeks prior to treatment with rusfertide is approximately 2.8 and in the 16 weeks post-treatment with rusfertide is approximately 0.17. Considering these average phlebotomies prior to treatment with rusfertide as placebo and post-rusfertide as on treatment and accounting for the extended duration of 32 weeks, it is assumed that for the double-blind treatment the mean number of phlebotomies to be approximately 3 and 1 for placebo and rusfertide over 32 weeks, respectively, with a standard deviation of 4. Based on these assumptions 106 subjects will be required per treatment group to detect a treatment difference in mean number of phlebotomies between the placebo and rusfertide treatment arms at a 0.05 level of significance (2-sided) with at least 95% power using a two-sample t-test.

As part of advice from the EMA, this endpoint will be presented as the primary efficacy endpoint for EMA submissions. The proportion of subjects achieving a response (absence of phlebotomy-eligibility) will be the first key secondary endpoint for the EMA submissions.

6.2.2. Key Secondary Endpoint #2: Proportion of Subjects with All Hematocrit Values <45% Between Baseline Through Week 32 (Inclusive)

It is assumed that approximately 60% of the subjects treated with rusfertide will have all hematocrit values <45% between Baseline through Week 32 (inclusive), with a single hematocrit value $\geq 45\%$ being allowed compared to 35% for placebo-treated subjects starting at Baseline and continuing through Week 32. Based on these response rate assumptions, 106 subjects will be

required per treatment group to detect a difference between rusfertide and placebo for this endpoint with at least 95% power, with a 5% 2-sided level of significance, using a z-test based on the large sample normal approximation to binomial distribution.

6.2.3. Key Secondary Endpoint #3: Mean Change from Baseline in Total Fatigue Score Based on PROMIS Short Form 8a at Week 32

Based on a survey of 538 subjects with PV ([Emanuel 2012](#)), fatigue carried the highest symptom intensity using the MPN-SAF 10 (mean = 4.4 with standard deviation of 2.8). Given the impact of this symptom on subjects' functionality the mean change from baseline in total fatigue score based on PROMIS Short Form 8a will be compared between rusfertide and placebo. Assuming a standard deviation of 2.8 units, a sample size of 125 in each group will have at least 80% power using a two-sample t-test to detect a difference of a 1-unit change from baseline with a 0.05 2-sided level of significance.

6.2.4. Key Secondary Endpoint #4: Mean Change from Baseline in Total Score Based on MFSAF v4.0 at Week 32

The Myelofibrosis Symptoms Assessment Form (MFSAF v4.0) is a 7-item questionnaire which allows for the assessment of patient-reported symptom manifestation ([Gwaltney 2017](#)). Assuming a standard deviation of 5.5 units a sample size of 125 in each group will have at least 80% power using a two-sample t-test to detect a difference of a 2-unit change from baseline with a 0.05 2-sided level of significance.

7. GENERAL ANALYSIS CONSIDERATIONS

7.1. General Rules

The statistical analyses will be reported using summary tables, listings, and figures (TLFs). The International Council for Harmonisation (ICH) numbering convention will be used for all TLFs.

Descriptive statistics will be provided for all efficacy, safety, PK and PD data. Continuous variables will be summarized by number of non-missing observations, means, standard deviations (SD), medians, minimums (Min) and maximums (Max). Additionally, for PK data, geometric mean and coefficient of variation will be presented. Categorical variables will be summarized by counts and by percentage of subjects in corresponding categories. In general, percentages for missing categories are not presented. Percentage calculations will be based on the total count of available data, if not otherwise mentioned. All confidence intervals will be presented at the nominal 95% level, unless otherwise stated.

Primary and key secondary efficacy analyses will be conducted after all subjects have completed Part 1a or discontinued Part 1a prematurely. To evaluate durability of response for rusfertide, analyses will be conducted after all subjects have completed Part 1b or discontinued Part 1b prematurely. Comprehensive efficacy and safety analyses will be conducted after all subjects in the study have either completed 156 weeks in the study or discontinued prematurely.

In general, when more than one numerical result is available at a particular timepoint, the average will be taken for analyses and will be identified as such in the listings. When more than one categorical result is available at a particular time point, the worst result will be taken for analyses (for example, if one result is “TRACE” and another result is “POSITIVE”, then the “POSITIVE” result is used for analysis) and will be identified as such in the listings.

For subjects who were rescreened because the first screening did not meet the study inclusion/exclusion criteria, data from the rescreened assessment will be used as the basis for Screening visit in the analyses.

If the time of first dosing is missing then the randomization time will be used as the imputed first dosing time, provided that the first dosing date is the same as the randomization date.

All analyses and tabulations will be performed using SAS[®] software Version 9.4 or higher. Pharmacokinetic and pharmacodynamic analyses may be performed using the SAS or Phoenix[®] platform (WinNonlin[®] and NLME) Version 8.3 or higher.

7.2. Specific Rules for Questionnaires

If more than one PGI-S, PGI-C, or EORTC questionnaire is completed on a particular day, the last one completed will be selected. For other questionnaires, the following rules will be used:

- For PROMIS, if there are two questionnaires done on the same day, the most complete one will be selected. If two of them are equally complete, the latest one will be selected. The PROMIS T-score will be obtained via the Item-Response Theory (IRT) response-pattern scoring approach using the PROMIS Scoring Service available at https://www.assessmentcenter.net/ac_scoringervice. The calibration sample is the default “PROMIS Wave 1” (Cella 2010).

- For MFSAF, if there are two questionnaires done on the same day, the most complete one will be selected. If two of them are equally complete, the latest one will be selected. The MFSAF total score will be generated only if none of the items are missing.
- For MPN-SAF, the total score is only calculated if 6 or more individual scores are available ([Emanuel 2012](#)).
- Questionnaires (except MFSAF) that are done within +/- 14 days of the scheduled visit will be used for that visit. However, for Week 32 dosing, only -14 days window will be allowed because subjects switch treatment from double-blind to open-label at Week 32 dosing.
- The EORTC QLQ-C30 is composed of both multi-item scales and single-item measures. Each of the multi-item scales includes a different set of items - no item occurs in more than one scale.

Table 1: EORTC Scales

	Abbreviation	Number of Items	Item Range ^a	Version 3.0 Item Numbers
Global health status/QoL				
Global health status/QoL (revised)	QL2	2	6	29, 30
Functional scales				
Physical functioning (revised)	PF2	5	3	1 to 5
Role functioning (revised)	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21 to 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
Symptom scales/items				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17
Financial difficulties	FI	1	3	28

Source: [Quality of Life Unit, EORTC Data Center 2001](#)

^a Item range is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving range = 3
QoL=quality of life.

The principle for scoring these scales is the same in all cases:

1. Estimate the average of the items that contribute to the scale; this is the raw score.
2. Use a linear transformation to standardize the raw score, so that scores range from 0 to 100; a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

The technical details are provided below ([Quality of Life Unit, EORTC Data Center 2001](#)):

If items $I_1, I_2, \dots, I_n$ are included in a scale, calculate the raw score (RS) as: $RS = (I_1 + I_2 + \dots + I_n)/n$

Apply the linear transformation as follows:

- Functional scales: Scale score = $\left\{1 - \frac{(RS-1)}{range}\right\} \times 100$
- Symptom scales / items: Scale score = $\left\{\frac{(RS-1)}{range}\right\} \times 100$
- Global health status/QoL: Scale score = $\left\{\frac{(RS-1)}{range}\right\} \times 100$

where range is the difference between the maximum possible value of RS and the minimum possible value. The QLQ-C30 has been designed so that all items in any scale take the same range of values. Therefore, the range of RS equals the range of the item values.

If at least half the components of a scale are present, then the scale score will be calculated using the average of all items answered as the raw score; otherwise the score will be set to missing. For single measures, if the item is missing the scale score is set to missing.

8. NOTATION OF DOSE LEVEL FOR STUDY DRUG AND VISITS

8.1. Notation of Dose Levels for Study Drug

The following notation will be used for the various dose levels of the study drug. “Two times a week” implies that the drug is administered with a gap of 3 or 4 days.

Table 2: Notation of Dose Levels for Study Drug

	Frequency	Dose	Notation Used Throughout Tables, Listings and Figures
Part 1a and 1b	Weekly (/wk)	10 mg 20 mg 30 mg 45 mg 60 mg 75 mg 90 mg	10 mg/wk 20 mg/wk 30 mg/wk 45 mg/wk 60 mg/wk 75 mg/wk 90 mg/wk
Part 2a	Weekly (/wk)	10 mg 20 mg 30 mg 40 mg 45 mg 60 mg 75 mg 90 mg	10 mg/wk 20 mg/wk 30 mg/wk 40 mg/wk 45 mg/wk 60 mg/wk 75 mg/wk 90 mg/wk
	Weekly alternating dose	10 mg/20 mg 20 mg/30 mg 30 mg/45 mg 45 mg/60 mg 60 mg/75 mg 75 mg/90 mg	10/20 mg/wk-Alt 20/30 mg/wk-Alt 30/45 mg/wk-Alt 45/60 mg/wk-Alt 60/75 mg/wk-Alt 75/90 mg/wk-Alt
	Two times a week dose	10 mg/10 mg 10 mg/20 mg 20 mg/20 mg 20 mg/30 mg 30 mg/30 mg 30 mg/45 mg 45 mg/45 mg 45 mg/60 mg 60 mg/60 mg	10/10 mg 2x/wk 10/20 mg 2x/wk 20/20 mg 2x/wk 20/30 mg 2x/wk 30/30 mg 2x/wk 30/45 mg 2x/wk 45/45 mg 2x/wk 45/60 mg 2x/wk 60/60 mg 2x/wk

	Frequency	Dose	Notation Used Throughout Tables, Listings and Figures
	Every two weeks dose	10 mg	10 mg/2wk
		20 mg	20 mg/2wk
		30 mg	30 mg/2wk
		40 mg	40 mg/2wk
		45 mg	45 mg/2wk
		50 mg	50 mg/2wk
		60 mg	60 mg/2wk
		75 mg	75 mg/2wk
		90 mg	90 mg/2wk

Note: -Alt: with “alternating dose level.”

8.2. Analysis Epochs

All visits are assigned to “epochs”:

- **Screening:**
The Screening epoch begins with the signing of informed consent and ends with the first dose of study drug.
- **Double-Blind Treatment:**
The blinded treatment epoch begins with the first dose of study drug and ends at the first dose of study drug of Part 1b.
- **Open-Label Treatment:**
The open-label treatment epoch begins with the first dose of study drug in Part 1b and ends at the first dose of study drug of Part 2.
- **Open-Label Extension Treatment:**
The open-label extension treatment epoch begins with the first dose of study drug in Part 2 and ends at the start of the Post-Treatment Safety Follow-Up phase. The Post-Treatment Safety Follow-up begins the day after the End of Treatment (EOT) visit.
- **Post-Treatment Safety Follow-up:**
The follow-up epoch begins after the EOT visit and ends at the End of Study (EOS) visit.

Vital signs, laboratory results and phlebotomies, if collected on a scheduled visit that marks a transition from one epoch to another are assigned to the previous epoch; e.g., vital signs collected on Baseline visit will be assigned to Screening epoch. For data collected with time (e.g., ECG, PK/PD parameters) on a scheduled visit that marks a transition from one epoch to another, if the time of the assessment is known and is after the dosing time, then the assessment will be assigned to the next epoch as appropriate.

For assignment of TEAEs to an epoch,

- AEs occurring on the first day of dosing will be identified as TEAEs and associated with the Double-Blind Treatment Epoch.
- If AE time of onset is available, then the date and time of onset will be used to assign AEs to an epoch as appropriate relative to the dosing date and time.
- If AE time of onset is not available,
 - AEs that are not injection site reactions occurring on the first day of dosing of the Open-Label Treatment Epoch or the Open-Label Extension Treatment Epoch will be associated with the previous Epoch (Double-Blind Treatment Epoch or Open-Label Treatment Epoch, respectively).
 - Injection site reactions occurring on the first day of dosing of the Open-Label Treatment Epoch or the Open-Label Extension Treatment Epoch will be associated with the current Epoch (Open-Label Treatment Epoch or Open-Label Extension Treatment Epoch, respectively).

For non-commercial use only

9. BASELINE, STUDY DAY, AND TIME WINDOWS

9.1. Baseline for Study

In general, baseline is defined as the last assessment done on the date prior to or the same as the date of the first dose of study drug, unless specified below. If the first dose date is not available, then the randomization date will be used. For height, an assessment at an unscheduled visit may be used as the baseline value if baseline assessment was not performed.

9.1.1. PK, PD, and ECG Parameters

For PK, PD, and ECG parameters, baseline is defined as the last assessment done prior to the first dose of study drug.

9.1.2. MFSAF

Average score of the 7-day interval immediately on or prior to first dose date. Specifically, the baseline MFSAF will be the average score of Day 1, Day -1, Day -2, Day -3, Day -4, Day -5, and Day -6 (Day 0 does not exist). If at least 4 of the 7 days are available, the average score will be calculated ([Gwaltney 2017](#)). If the baseline is not derivable using the above rule, then the screening record, if available using the above rule, will be used for the baseline.

9.2. Part 1b: Week 32

For the purpose of establishing an additional anchor for subjects switching from placebo to rusfertide after Part 1a, a “Part 1b: Week 32 Anchor” is defined as the last assessment done on the date prior to or the same as the date of the first dose of study drug in Part 1b, unless specified below.

9.2.1. MFSAF

Average score of the 7-day interval immediately on or prior to first dose date in Part 1b. For example, if the first dose date for Part 1b is on Day 225, then the baseline MFSAF for Part 1b will be the average score of Day 225, Day 224, Day 223, Day 222, Day 221, Day 220, and Day 219. If at least 4 of the 7 days are available, the average score will be calculated.

9.3. Part 2: Week 52

Similarly, a “Part 2: Week 52 Anchor” is defined as the last assessment done on the date prior to or the same as the date of the first dose of study drug in Part 2, unless specified below.

9.3.1. MFSAF

Average score of the 7-day interval immediately on or prior to first dose date in Part 2. For example, if the first dose date for Part 2 is on Day 365, then the baseline MFSAF will be the average score of Day 365, Day 364, Day 363, Day 362, Day 361, Day 360, and Day 359. If at least 4 of the 7 days are available, the average score will be calculated.

9.4. Change from Baseline

- Change from baseline = postbaseline result – baseline result.

- %Change from baseline = (change from baseline/baseline result) x 100.

9.5. Study Day

Study day is calculated from the dosing date of study drug.

For postdose study day: Assessment date – First Dosing Date + 1

For predose study day: Assessment date – First Dosing Date

Study Day 1 is the day of first study drug dosing.

9.6. Analysis Visits

Protocol-Scheduled Visits:

Except for baseline and early termination (ET), assessments done at protocol-scheduled visits will be analyzed based on the actual visits performed, except for assessments done in the Week 32 visit folder (see below). Assessments attributed to an unscheduled visit but occurring on the same date as a scheduled visit will be mapped to and relabeled as the relevant scheduled visit. For by-visit summaries, results of assessments at unscheduled visits on dates that do not coincide with scheduled visits will be listed but not analyzed. For analyses that are inclusive of a period (e.g., proportion of subjects with all hematocrit (HCT) < 45% between Baseline through Week 32, or time to first HCT ≥ 45%), all scheduled and unscheduled assessments will be considered.

Analysis baseline visits will be created according to baseline specifications (see [Section 9.1](#)).

For assessments done in the Week 32 visit folder, if the assessment dates are on or before the first dose date of Part 1b, the latest assessments will be used for the Week 32 analyses. If the assessment dates are after the first dose date of Part 1b, the assessments will be excluded from Week 32 analyses.

For those subjects who withdraw early from the study,

- If all doses between the last scheduled visit and the Early Termination visit are administered, the results at the actual Early Termination visit will be used for the analysis Early Termination visit. In addition, a copy of the Early Termination visit will be created and assigned the next scheduled visit.
- If the dose administration between the last scheduled visit and the Early Termination visit is not complete or the Early Termination visit occurs after the next planned scheduled visit, then a copy of the last scheduled visit will be created and used for the analysis of Early Termination visit. The results at the actual Early Termination visit will not be analyzed unless needed for Safety Follow-up analysis.

All subjects will have an analysis Safety Follow-up visit. This visit will correspond to the actual Safety Follow-up visit 4-weeks post last dose, but if a Safety Follow-up visit does not exist, the analysis Safety Follow-up visit will include the last available visit after End of Treatment. If the last visit is the End of Treatment visit, then it will not be copied and reassigned to the Safety Follow-up visit.

10. ANALYSIS POPULATIONS

The following populations will be used:

10.1. Intent-to-Treat Set (ITT)

All randomized subjects according to the treatment assigned at randomization. For the US submission, the ITT will be the primary analysis population for efficacy endpoints.

10.2. Full Analysis Set (FAS)

All randomized subjects according to the treatment assigned at randomization (ITT principle), receive at least a dose of study drug, have a baseline and at least one postbaseline assessment in Part 1a. For the EMA submission, the FAS will be the primary analysis population for efficacy endpoints.

10.3. Safety Analysis Set (SAS)

All subjects who receive at least one dose of study drug. This population will be used for all safety analyses. Subjects will be analyzed according to the study treatment they actually receive. Subjects who receive at least one dose of rusfertide will be summarized in the rusfertide group, regardless of the original treatment assignment. For safety analysis by subgroup of stratification factors, the actual stratification will be used.

10.4. Per Protocol Set (PPS)

Subset of subjects in ITT with no major protocol deviations at study entry or during the study that could impact efficacy endpoints.

The determination of the inclusion into PPS will be done before database soft lock for unblinding. This set will be used for supportive analyses of the primary and key secondary endpoints.

10.5. PK Analysis Set (PKAS)

Subset of subjects in ITT who have received at least one dose of study drug and have at least one measurable postbaseline PK concentration.

10.6. PD Analysis Set (PDAS)

Subset of subjects in ITT who have received at least one dose of study drug and have at least one postbaseline PD assessment.

11. DESCRIPTION OF STUDY POPULATION

11.1. Subject Disposition

Subject disposition at the end of study will be summarized for all subjects for Part 1a (by treatment), Part 1b and Part 2 and overall. Summaries will include: the number of screened subjects, the number of screen failures, the number of randomized subjects, the number of subjects who were treated with study drug, the number of subjects in each analysis set (intent-to-treat set, full analysis set, safety analysis set, per protocol analysis set, completers analysis set, PK analysis set, and PD analysis set), the number of subjects entering and completing each part of the study, the number of subjects terminating the treatment early. The primary reasons for early discontinuation from treatment will be tabulated. The subject disposition at the end of study, 4 weeks after last dose (post-treatment safety follow-up) will be presented in listings. All disposition data will also be listed.

11.2. Enrollment

An enrollment summary by site, country, and region will be tabulated for the ITT.

11.3. Demographics and Baseline Characteristics

The demographics and baseline characteristics will be summarized using descriptive statistics and frequency distributions using the ITT and Safety Analysis Set. Demographics and baseline characteristics data will also be included in listings.

Demographics information includes age (as entered in the Electronic Data Capture [EDC]), sex, race, ethnicity, weight, height, and body mass index (BMI). BMI is calculated as follows:

$$\text{BMI (kg/m}^2\text{)} = \text{Body Weight (kg)} / (\text{Height}^2 \text{ (m)}); \text{ rounded to one decimal place}$$

Baseline disease characteristics include:

- Eastern Cooperative Oncology Group (ECOG) classification
- Age at PV diagnosis
- Duration of PV:
 - If PV duration based on PV diagnosis date provided in the EDC < 1 year, the PV onset date will be defined as the earliest date entered in the EDC of the following 4 dates (if a date is completely missing, then impute to Informed Consent date):
 - PV Diagnosis date (if PV diagnosis date is incomplete, then impute Jan 01)
 - First HCT date where HCT $\geq$ 45%
 - First available phlebotomy date (if the first available phlebotomy date is incomplete, then impute July 15)
 - Start date of first cytoreductive therapy
 - If PV duration based on PV diagnosis data provided in the EDC $\geq$ 1 year, then PV onset date = PV diagnosis date

- The duration of PV = (Screening date – PV onset date)
- Phlebotomies prior to rusfertide (number, rate/year, and average time between phlebotomies)
- Hematocrit, hemoglobin, white blood cells (WBCs), red blood cells (RBCs), platelets, ferritin, TSAT, serum iron, JAK2 mutation burden
- Prior and concurrent cytoreductive therapy
- Prior thromboembolic event status
- Risk category: High (age ≥ 60 years or prior thromboembolic event), Low (age < 60 years and no prior thromboembolic event)
- Cancer history prior to rusfertide
- Baseline spleen size (cm) by palpation and by magnetic resonance imaging (MRI)/computed tomography(CT) scan/ultrasound
- Baseline PROMIS Fatigue Short Form 8a, MFSAF TSS4, MFSAF TSS7, and MPN-SAF, PGI-S, and EORTC QLQ-C30

11.4. Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The number and percentage of subjects with any medical history will be summarized in a table organized by treatment, and overall by each system organ class (SOC) and preferred term (PT) using the ITT and Safety Analysis Set. Percentages will be calculated based on the number of subjects in the ITT and Safety Analysis Set, respectively. Subject medical history data including specific details will be presented in a listing organized by subject ID that includes a description of the Medical Condition/Event, Onset Date, Ongoing, End Date, and CTCAE Grade for the ITT and Safety Analysis Set. All medical history data will be included in listings.

Events that are entered in the Medical History case report form will be considered medical history events regardless of the onset and resolution dates.

Cancer events are selected using the FDA Medical Queries (FMQ) – narrow terms for malignancy-excluding preferred term myeloproliferative neoplasm. Thromboembolic events are selected using FMQ – narrow terms for thrombosis (<https://www.fda.gov/media/164639/download>).

11.5. Physical Examination

All physical examination results with any clinically significant abnormal findings will be included in data listing for the Safety Analysis Set. Physical examination data collected at unscheduled visits will not be summarized but will be included in data listings.

11.6. Protocol Deviations

Major protocol deviations that could potentially affect the efficacy or safety conclusions of the study will be identified prior to database lock and unblinding. These deviations may include, but are not limited to:

- Randomized subjects who do not satisfy selected inclusion and exclusion criteria
- Randomized subjects who are mis-stratified
- Randomized subjects who receive the wrong treatment
- Subjects who receive an excluded concomitant treatment

A listing of all protocol deviations including the deviation designation (e.g., Major), category, and indication of whether the deviation led to an exclusion of a subject will be presented in a data listing for the ITT.

Major protocol deviations will be summarized by deviation category using the ITT.

For non-commercial use only

12. EFFICACY ANALYSES

All efficacy analyses will be performed using ITT for the US submission and using the FAS for the EMA submission.

12.1. General Consideration

- To determine the historical rate of phlebotomies the following rules are implemented:
 - Phlebotomies performed within an 8-day period will be counted as a single phlebotomy.
 - When two phlebotomies are within an 8-day period and are already counted as a single phlebotomy, then the next phlebotomy will be counted as a new phlebotomy. For example, if 3 phlebotomies occur on Day 1, Day 7, and Day 13, then the first two phlebotomies will be counted as a single phlebotomy, but the third phlebotomy will be counted as another phlebotomy even though it is within 7 days of the 2nd phlebotomy.
 - If a subject has a phlebotomy on the first day of dosing it is counted towards the historical rate.
 - The interval during which the phlebotomies were received prior to receiving rusfertide is estimated as

$$\text{Duration (weeks)} = (\text{TRTSDT} - \text{MINDT} + \text{MEDPHBTW})/7$$

TRTSDT = Date of first exposure to rusfertide

MINDT = Date of earliest available historical date for phlebotomy prior to rusfertide exposure

MEDPHBTW = Median number of days between phlebotomies based on available historical dates

- For End of Part visits, a window of 3 days will be added to the End of Part visit date to allow for phlebotomies occurring right after the End of Part visit.
- In general, complete blood count values at local labs will be used for analyses, unless otherwise specified.
- For stratified analyses of primary and key secondary efficacy endpoints, the stratification factor will be as randomized in IxRS. Sensitivity analyses will be performed using the stratification factor entered in EDC. For subgroup analyses of safety related to the stratification information, the information will be the stratification factors entered in EDC.
- For stratified analyses of primary and key secondary efficacy endpoints, if some stratum is too small to facilitate the analyses (e.g., the model fails to converge), the smallest stratum will be combined with the next smallest stratum for analyses.
- If a subject meets phlebotomy-eligible by having confirmed HCT $\geq 45\%$ and that is at least 3% higher than the baseline hematocrit, the first occurrence (not the confirmed value) will be used for time to phlebotomy-eligible.

- Time to event analyses during Part 1a (e.g., time to first hematocrit $\geq 45\%$ after randomization):
 - Subjects will be censored at the end of Part 1a (or at early termination) if they have not experienced the event.
 - The start of the time will be the first dose of study drug for subjects who were dosed with study drug. The start of the time will be the randomization for subjects who were randomized but not dosed with study drug.
 - The end of the time will be the time of the event or censored time.
- Time to event analyses during Part 1b (e.g., time to first hematocrit $\geq 45\%$ during Part 1b excluding Part 1a):
 - Only subjects who are eligible at the start of Part 1b are included (e.g., subjects with HCT $< 45\%$ at the start of Part 1b are eligible for time to first HCT $\geq 45\%$).
 - Subjects will be censored at the end of Part 1b (or at early termination) if they have not experienced the event.
 - The start of the time will be the first dose of study drug of Part 1b.
 - The end of the time will be the time of the event or censored time.
 - Subjects who discontinue prior to Part 1b are excluded.

12.2. Analysis of Efficacy Endpoints

12.2.1. Primary Efficacy Endpoint

The proportion of subjects achieving a response starting at Week 20 through Week 32 (inclusive) is the primary endpoint to evaluate the efficacy of rusfertide compared to placebo. A response is defined as absence of phlebotomy eligibility starting at Week 20 through Week 32 (inclusive). Phlebotomy eligibility is defined as a confirmed hematocrit $\geq 45\%$ and that is at least 3% higher than the baseline hematocrit **or** a hematocrit $\geq 48\%$. Confirmation is defined as 2 consecutive hematocrit assessments that are $\geq 45\%$ and at least 3% higher than the baseline hematocrit. Subjects will also be classified as non-responders according to [Section 12.3](#). Hematocrit values from each clinical site's local laboratory will be used for the primary endpoint.

The ITT will be used for the primary efficacy analysis. Intercurrent events are described in [Section 12.3](#).

The primary efficacy endpoint of difference in proportions between rusfertide and placebo will be analyzed using Cochran-Mantel-Haenszel (CMH) (General Association statistic) test at a 0.05 2-sided level of significance, stratified for ongoing PV therapy at randomization. The stratification variable will be according to the assignment in the IxRS at randomization.

If a subject becomes phlebotomy eligible prior to Week 20 but maintains absence of phlebotomy eligibility between Week 20 and Week 32 (both inclusive) the subject will be considered as a responder. Subjects who will also be classified as non-responders are defined in [Section 12.3](#). The Mantel-Haenszel stratum-adjusted, common risk difference and 95% confidence interval (CI) will be used as a measure of the difference in treatment response. In addition, the overall

stratum-adjusted odds ratio (and 95% CI) will be presented as a measure of the association between treatment and response.

As part of the advice from the EMA, it is proposed the first key secondary endpoint described in [Section 12.2.2.1](#), average number of phlebotomies between Baseline through Week 32 (inclusive) be presented as the primary efficacy endpoint for the EMA submissions. The proportion of subjects achieving a response (absence of phlebotomy-eligibility) between Week 20 and Week 32 (inclusive) will be the first key secondary endpoint for the EMA submissions.

12.2.2. Key Secondary Efficacy Endpoints

The four key secondary endpoints outlined below will be tested for treatment differences through a closed testing procedure in a single hierarchical order as described in [Section 12.5](#). The ITT will be used for the key secondary efficacy analyses.

12.2.2.1. Key Secondary Endpoint #1: Mean Number of Phlebotomies Between Baseline Through Week 32 (Inclusive)

The mean number of phlebotomies between Baseline through Week 32 (inclusive) will be analyzed using an analysis of covariance (ANCOVA). The model will include the number of phlebotomies in 32 weeks prior to study treatment as a covariate and fixed effects of treatment, and stratification variable (ongoing PV therapy at randomization). Intercurrent events are described in [Section 12.3](#). The comparison will be based on the difference in least squares mean between treatments.

This mean number of phlebotomies between Baseline through Week 32 endpoint will be used as the primary efficacy endpoint for EMA submissions. The proportion of subjects achieving a response (absence of phlebotomy-eligibility) will be the first key secondary endpoint for the EMA submissions.

12.2.2.2. Key Secondary Endpoint #2: Proportion of Subjects with All Hematocrit Values <45% Between Baseline Through Week 32 (Inclusive)

The proportion of subjects with all hematocrit values <45% between Baseline through Week 32 (inclusive) will be analyzed similar to the primary endpoint using CMH 2-sided test at a 0.05 level of significance, stratified for ongoing PV therapy at randomization. A single hematocrit value $\geq 45\%$ between Baseline and Week 32 will be allowed in this analysis. Intercurrent events are described in [Section 12.3](#).

12.2.2.3. Key Secondary Endpoint #3: Mean Change from Baseline at End of Part 1a (Week 32) in the PROMIS Fatigue Short Form 8a Total T-score

The PROMIS Fatigue SF 8a T-score will be obtained via the IRT response-pattern scoring approach using the PROMIS Scoring Service available at https://www.assessmentcenter.net/ac_scoring-service. The calibration sample is the default “PROMIS Wave 1” ([Cella 2010](#)).

The mean change from baseline in T-score using PROMIS Fatigue SF 8a will be analyzed using a mixed model for repeated measures (MMRM) and will include treatment, scheduled visit, and

stratification variable (ongoing PV therapy at randomization) as fixed factors, treatment-by-visit interaction, baseline-by-visit interaction, and baseline PROMIS Fatigue SF 8a T-score as a covariate. The model will use an unstructured within subject covariance matrix and a restricted maximum likelihood (ReML) estimation method. The degree of freedom of the denominator will be estimated using the Kenward-Roger method. The primary contrast of interest is the least squares mean difference between rusfertide and placebo at Week 32. Intercurrent events are described in [Section 12.3](#).

If the model fails to converge using an unstructured covariance matrix, a simpler covariance structure (e.g., heterogeneous Toeplitz [TOEPH], heterogeneous autoregressive of order 1 [ARH(1)] and autoregressive of order 1 [AR(1)], in that order) may be employed to achieve convergence.

12.2.2.4. Key Secondary Endpoint #4: Mean Change from Baseline at End of Part 1a (Week 32) in the MFSAF TSS7

The MFSAF TSS7 is the sum of seven items in the MFSAF. The weekly total symptom score will be calculated by averaging daily scores over a 7-day interval leading to Week 32 (inclusive).

The mean change from baseline in MFSAF TSS7 will be analyzed using an MMRM model that will include treatment, scheduled visit, and stratification variable (ongoing PV therapy at randomization) as fixed factors, treatment-by-visit interaction, baseline-by-visit interaction, and baseline MFSAF TSS7 as a covariate. The model will use an unstructured within subject covariance matrix and a restricted maximum likelihood (ReML) estimation method. The degree of freedom of the denominator will be estimated using the Kenward-Roger method. The weekly total symptom score at each scheduled visit will be calculated by averaging the daily scores over a 7-day interval leading to that scheduled visit. The primary contrast of interest is the least squares mean difference between rusfertide and placebo at Week 32. Intercurrent events are described in [Section 12.3](#).

If the model fails to converge using an unstructured covariance matrix, a simpler covariance structure (e.g., TOEPH, ARH(1) and AR(1), in that order) may be employed to achieve convergence.

12.2.3. Analysis of Other Secondary Endpoints

12.2.3.1. Analysis of PV-focused PRO-Related Secondary Endpoints in Subjects with Moderate or Severe Symptoms at Baseline

12.2.3.1.1. Mean Change from Baseline at end of Part 1a (Week 32) in the PROMIS Fatigue Short Form 8a total T-score among Patients Who Have Moderate or Severe Symptoms at Baseline

The PROMIS Fatigue SF 8a T-score will be obtained via the IRT response-pattern scoring approach using the PROMIS Scoring Service available at https://www.assessmentcenter.net/ac_scoringservice. The calibration sample is the default “PROMIS Wave 1” ([Cella 2010](#)). Moderate or severe symptoms are defined as a baseline PROMIS Fatigue SF 8a T-score ≥ 58.8 based on descriptive analysis approach described in Psychometric Analysis Plan (PAP) v2.0. In the descriptive analysis approach the distribution of

the PROMIS Fatigue SF 8a total T-score at baseline was categorized using the PGI-S. The response quartiles were used to establish the recommended cut-off points for the PROMIS Fatigue SF 8a total T-score corresponding to “none” or “mild” vs “moderate” or “severe” categories on the PGI-S (See [Appendix 3](#)).

The mean change from baseline in T-score using PROMIS Fatigue SF 8a among patients who have moderate or severe symptoms at baseline will be analyzed using a mixed model for repeated measures (MMRM) and will include treatment, scheduled visit, and stratification variable (ongoing PV therapy at randomization) as fixed factors, treatment-by-visit interaction, baseline-by-visit interaction, and baseline PROMIS Fatigue SF 8a T-score as a covariate. The model will use an unstructured within subject covariance matrix and a restricted maximum likelihood (ReML) estimation method. The degree of freedom of the denominator will be estimated using the Kenward-Roger method. The primary contrast of interest is the least squares mean difference between rusfertide and placebo at Week 32.

12.2.3.1.2. Mean Change from Baseline at end of Part 1a (Week 32) in the MFSAF TSS4 Among Patients Who Have Moderate or Severe Symptoms at Baseline

The MFSAF TSS4 is the sum of four items in the MFSAF (fatigue, itching [pruritus], night sweats, and abdominal discomfort). The weekly total symptom score will be calculated by averaging daily scores over a 7-day interval leading to Week 32 (inclusive). Moderate or severe symptoms are defined as a baseline MFSAF TSS4 ≥ 9.1 based on descriptive analysis approach described in PAP v2.0. In the descriptive analysis approach the distribution of the MFSAF TSS4 at baseline was categorized using the PGI-S. The response quartiles were used to establish the recommended cut-off points for the MFSAF TSS4 corresponding to “none” or “mild” vs “moderate or “severe” categories on the PGI-S (See [Appendix 3](#)).

The mean change from baseline in MFSAF TSS4 among patients who have moderate or severe symptoms at baseline will be analyzed using an MMRM model that will include treatment, scheduled visit, and stratification variable (ongoing PV therapy at randomization) as fixed factors, treatment-by-visit interaction, baseline-by-visit interaction, and baseline MFSAF TSS4 as a covariate. The model will use an unstructured within subject covariance matrix and a ReML estimation method. The degree of freedom of the denominator will be estimated using the Kenward-Roger method. The weekly total symptom score at each scheduled visit will be calculated by averaging the daily scores over a 7-day interval leading to that scheduled visit. The primary contrast of interest is the least squares mean difference between rusfertide and placebo at Week 32.

12.2.3.1.3. Mean Change from Baseline at end of Part 1a (Week 32) in the MPN-SAF Problems with Concentration among Patients Who Have Moderate or Severe Symptoms at Baseline

Moderate or severe symptoms are defined as a baseline MPN-SAF problems with concentration ≥ 4 ([Emanuel 2012](#)). This definition of moderate or severe symptoms was also confirmed based on descriptive analysis approach described in PAP v2.0. In the descriptive analysis approach the distribution of the MPN-SAF problems with concentration at baseline was categorized using the PGI-S. The response quartiles were used to establish the recommended cut-off points for the

MPN-SAF Problems with Concentration corresponding to “none” or “mild” vs “moderate or severe” categories on the PGI-S (See [Appendix 3](#)).

The mean change from baseline in MPN-SAF problems with concentration among patients who have moderate or severe symptoms at baseline will be analyzed using an MMRM model that will include treatment, scheduled visit, and stratification variable (ongoing PV therapy at randomization) as fixed factors, treatment-by-visit interaction, baseline-by-visit interaction, and baseline MPN-SAF problems with concentration as a covariate. The model will use an unstructured within subject covariance matrix and a ReML estimation method. The degree of freedom of the denominator will be estimated using the Kenward-Roger method. The primary contrast of interest is the least squares mean difference between rusfertide and placebo at Week 32.

12.2.3.1.4. Proportion of Subjects With a Meaningful Within-Person Change at Week 32 in the PROMIS Fatigue SF 8a T-score, MFSAF TSS4, and MPN-SAF Problems With Concentration Among Subjects Who Have Moderate or Severe Symptoms at Baseline

The proportion of subjects with a meaningful within-person change at Week 32 (inclusive) in the PROMIS Fatigue SF 8a T-score, MFSAF TSS4, and MPN-SAF problems with concentration among subjects who have moderate or severe symptoms at baseline will be analyzed similar to the primary endpoint using CMH 2-sided test at a 0.05 level of significance, stratified for ongoing PV therapy. Subjects who discontinued early will be classified as not having achieved the meaningful within-person change at Week 32.

12.2.3.2. Analysis of Other PRO-related Secondary Endpoints:

All other categorical secondary efficacy endpoints will be analyzed using CMH test at a 0.05 2-sided level of significance, with adjustment for stratification factor of ongoing PV therapy. Subjects who discontinue early or have missing data at a scheduled assessment time point will be considered non-responders for that endpoint.

Continuous endpoints such as average of number of phlebotomies will be analyzed using ANCOVA with the number of phlebotomies in 32 weeks prior to study treatment as a covariate and treatment and stratification variable (ongoing PV therapy at randomization) as main effects.

Changes from baseline will be analyzed using MMRM. The model will include the fixed effects of treatment, scheduled visit, and stratification variable (ongoing PV therapy at randomization), treatment-by-visit interaction, baseline-by-visit interaction, and baseline scores as covariate. The model will use an unstructured within subject covariance matrix and a ReML estimation method. The degree of freedom of the denominator will be estimated using the Kenward-Roger method. If the model fails to converge using an unstructured covariance matrix, a simpler covariance structure (e.g., TOEPH, ARH(1) and AR(1), in that order) may be employed to achieve convergence. The comparison will be based on the difference in least squares mean (LSMEAN) between treatments at the specific time point.

12.2.3.3. Analysis of Hematocrit and Phlebotomy-related Secondary Endpoints:

All other categorical secondary efficacy endpoints will be analyzed using CMH test at a 0.05 2-sided level of significance, with adjustment for stratification factor of ongoing PV therapy.

Subjects who discontinue early or have missing data at a scheduled assessment time point will be considered non-responders for that endpoint.

To estimate the time to event such as phlebotomy or hematocrit value $\geq 45\%$, Kaplan-Meier estimate will be presented, and a stratified log-rank test will be used to compare treatment arms. The stratification variable is the ongoing PV therapy at randomization.

These analyses are exploratory in nature. No multiplicity adjustments will be made for other secondary endpoints.

Part 1a - Baseline through Week 32 (inclusive):

Comparison of rusfertide to placebo during randomized, double-blind, placebo-controlled, add-on parallel phase.

PV-focused PRO-related secondary endpoints in subjects with moderate or severe symptoms at baseline:

- Mean change from baseline at end of Part 1a (Week 32) in the Patient Reported Outcomes Measurement Information System (PROMIS) Fatigue Short Form 8a total T-score among patients who have moderate or severe symptoms at baseline.
- Mean change from baseline at end of Part 1a (Week 32) in the Myelofibrosis Symptom Assessment Form version 4.0 (MFSAF) Total Symptom Score-4 items (TSS4) among patients who have moderate or severe symptoms at baseline. TSS4 is the sum of 4 items of the MFSAF (fatigue, itching, night sweats, and abdominal discomfort). The weekly TSS4 at Week 32 will be calculated by averaging daily TSS4 over a 7-day interval leading to Week 32 (inclusive).
- Mean change from baseline at end of Part 1a (Week 32) in the Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF) Problems with Concentration among patients who have moderate or severe symptoms at baseline.
- Proportion of subjects with a meaningful within-person change at Week 32 in the PROMIS Fatigue SF8a T-score, MFSAF TSS4, and MPN-SAF problems with concentration among subjects who have moderate or severe symptoms at Baseline.

Other PRO-related secondary endpoints:

- Mean change from baseline at each scheduled assessment in individual questions and the total fatigue T-score based on PROMIS® Short Form 8a for all ITT subjects.
- Mean change from baseline at each scheduled assessment in individual questions and the total fatigue T-score based on PROMIS® Short Form 8a for subjects who have moderate or severe symptoms at baseline.
- Mean change from baseline at each scheduled visit in individual symptom score and the MFSAF v4.0 TSS4 for all ITT subjects.
- Mean change from baseline at each scheduled visit in individual symptom score and the MFSAF v4.0 TSS4 for subjects who have moderate or severe symptoms at baseline.

- Proportion of subjects with a meaningful within-person change at Week 32 in the PROMIS Fatigue SF8a, MFSAF TSS4, and MPN-SAF problems with concentration among all ITT subjects.
- Mean change from baseline at each scheduled assessment in individual symptom score and the total score based on MPN-SAF TSS 10.
- Mean change from baseline in PGI-S at each scheduled assessment.
- Proportion of subjects with at least one unit improvement in the PGI-S, i.e., a ≥ 1 point decrease in the PGI-S, e.g., a change from “Moderate (2)” to “Mild (1)” or “Mild (1)” to “None (0)” at each scheduled visit.
- Proportion of subjects with a rating on the PGI-C of either “much better” or “a little better” at each scheduled assessment.
- Mean PGI-C at each scheduled assessment.
- Mean change from baseline at each scheduled visit in individual symptom score and the TSS7 based on MFSAF v4.0.
- Mean change from baseline in the individual domain scores and the total EORTC QLQ-C30 score at Week 32.

Hematocrit and Phlebotomy-related secondary endpoints:

- Median time to first hematocrit $\geq 45\%$ after randomization.
- Median time to phlebotomy-eligible after randomization.
- Proportion of subjects with at least one hematocrit $\geq 48\%$.
- Proportion of subjects who maintain absence of phlebotomy eligibility.
- Phlebotomy annualized rate.

Parts 1a+1b: Baseline through Week 52

- Proportion of subjects originally randomized to rusfertide achieving absence of phlebotomy eligibility (durable response) for 52 weeks.
- Median time to first hematocrit $\geq 45\%$ during Part 1b, excluding Part 1a. In this analysis, the time is reset at the first dose date of Part 1b. Only subjects who have HCT $< 45\%$ at the start of Part 1b are eligible for this analysis.
- Median time to first phlebotomy during Part 1b, excluding Part 1a. In this analysis, the time is reset at the first dose date of Part 1b.
- Proportion of subjects with all hematocrit values $< 45\%$ between Baseline through Week 52 (inclusive). Two hematocrit values $\geq 45\%$ are allowed. Hematocrit values from each clinical site’s local laboratory will be used for this endpoint.
- Proportion of subjects originally randomized to placebo who meet the phlebotomy eligibility in Part 1a and Part 1b, separately.

Parts 1a+1b+2: Baseline Through Week 156

The main goal of Part 2 is to evaluate maintenance of effect and safety during the extended period of treatment until Week 156. No hypothesis testing will be performed for Part 2. Endpoints described above will be presented as appropriate. Point estimates and the confidence intervals associated with the estimates may be provided.

12.3. Handling of Dropouts, Missing Data, Intercurrent Events, and Non-Responder Categories

The primary efficacy analysis will be conducted on the ITT.

The estimand framework of this study will use the treatment policy strategy, hypothetical strategy, and composite strategy. The treatment is rusfertide versus placebo. The population is the adult patients with polycythemia satisfying the inclusion/exclusion criteria.

The primary variable is the proportion of subjects achieving a response starting at Week 20 through Week 32 (inclusive). The population level summary is the treatment difference in the proportion of subjects achieving a response starting at Week 20 through Week 32 (inclusive).

The following events are defined as intercurrent events for the primary endpoint:

- Subject is missing hematocrit assessment at Week 32. The subject will be considered as a non-responder.
- Subject discontinues the study treatment prematurely during Part 1a for any reason. The subject will be considered as a non-responder.
- Subject has more than 1 missing hematocrit value at any scheduled visit between Week 20 and Week 32 (inclusive). The hematocrit value may be obtained within ± 1 week of the scheduled visit. The subject will be considered as a non-responder.
- Subject is phlebotomized between Week 20 and Week 32 (inclusive). The subject will be considered as a non-responder.
- Subject with $HCT \geq 45\%$ and that is at least 3% higher than the baseline hematocrit at Week 32 (i.e., no confirmation is required at Week 32). The subject will be considered as a non-responder.
- Subject receives prohibited medications or has prohibited changes in cytoreductive therapy between Baseline up to Week 32. The subject will be considered as a non-responder. As a sensitivity analysis, the subject's data will be used as is.
- Subject doesn't comply with the treatment. The subject's data will be used as is.
- Subject randomized but doesn't receive any treatment. The subject will be considered as a non-responder.
- Subject receives the wrong treatment according to the randomization. The subject's data will be used as randomized based on the intent-to-treat principle.
- Subject is mis-stratified. The subject's data will be used as randomized based on the intent-to-treat principle.

For the key secondary endpoints:

The same estimand framework as described for the primary endpoint will apply to the key secondary endpoints.

1. Average number of phlebotomies between Baseline and Week 32 (inclusive):

The following events are defined as intercurrent events:

- Subject discontinues the study treatment prematurely for any reason during Part 1a. The number of phlebotomies will be imputed using the historical rate as follows.

Phlebotomy data available prior to study treatment will be utilized for subjects who discontinue Part 1a earlier. It will be assumed that a subject continues to observe the same number of phlebotomies, i.e., no treatment benefit after the subject discontinues from the study. Based on this assumption, the number of phlebotomies prior to study treatment will be normalized to 32 weeks and will be imputed under the assumption that the distribution of missing values from the time of discontinuation through Week 32 (Day 224) is the same as that of observed values prior to randomization. For instance, if a subject has 12 weeks of observed data (i.e., early termination after Week 12 visit), then the number of phlebotomies will be (number of phlebotomies observed in the first 12 weeks + 20 weeks × rate/week in the pretreatment period).
- Subject receives prohibited medications or prohibited changes in cytoreductive therapy between Baseline up to Week 32. The subject's data after the use of prohibited medications or prohibited changes in cytoreductive therapy will be considered missing and will be imputed using the historical rates as above. As a sensitivity analysis, the subject's data will be used as is.
- Subject doesn't comply with the treatment. The subject's data will be used as is.
- Subject randomized but doesn't receive any treatment. The number of phlebotomies will be imputed by the historical data as above.
- Subject receives the wrong treatment according to the randomization. The subject's data will be used as randomized based on the intent-to-treat principle.
- Subject is mis-stratified. The subject's data will be used as randomized based on the intent-to-treat principle.

2. Proportion of subjects with all hematocrit values between Baseline through Week 32 (inclusive) $\geq 45\%$. A single hematocrit value $\geq 45\%$ is allowed:

The following events are defined as intercurrent events:

- Subject is missing a hematocrit assessment at Week 32 and had at least one value $\geq 45\%$ prior to Week 32. The subject will be considered as a non-responder.
- Subject discontinues the study treatment prematurely during Part 1a for any reason. The subject will be considered as a non-responder.
- Subject has more than two missing hematocrit values at any scheduled visit between Baseline and Week 32 (inclusive). The hematocrit value may be obtained within ± 1 week of the scheduled visit. The subject will be considered as a non-responder.

- Subject is phlebotomized between Baseline and Week 32 (inclusive). The subject will be considered as a non-responder.
 - Subject receives prohibited medications or prohibited changes in cytoreductive therapy between Baseline up to Week 32. The subject will be considered as a non-responder. As a sensitivity analysis, the subject's data will be used as is.
 - Subject doesn't comply with the treatment. The subject's data will be used as provided.
 - Subject randomized but doesn't receive any treatment. The subject will be considered as a non-responder.
 - Subject receives the wrong treatment according to the randomization. The subject's data will be used as provided based on the intent-to-treat principle.
 - Subject is mis-stratified. The subject's data will be used as randomized based on the intent-to-treat principle.
3. Mean change from baseline in the total fatigue T-score based on the PROMIS Fatigue Short Form 8a at Week 32:

The following events are defined as intercurrent events:

- Subject discontinues the study treatment prematurely during Part 1a for any reason. The subject's data will be used as provided. An MMRM approach will be employed to use all available visit data, including subjects with some missing visit data in order to arrive at an estimate of the mean treatment effect.
- Subject receives prohibited medications or prohibited changes in cytoreductive therapy between Baseline up to Week 32. The subject's data after the use of prohibited medications or prohibited changes in cytoreductive therapy will be considered missing. As a sensitivity analysis, the subject's data will be used as is.
- Subject doesn't comply with the treatment. The subject's data will be used as provided.
- Subject randomized but doesn't receive any treatment. The subject's data will be imputed as follows:
 - For each subject at each scheduled assessment (Baseline to Week 32), a simple random sample with replacement is drawn from all the remaining placebo subjects with data at that scheduled assessment. The seed will be 7707.
 - Repeat the step 100 times.
 - The score of each subject is replaced by the mean score of the 100 samples drawn.
- Subject receives the wrong treatment according to the randomization. The subject's data will be used as provided based on the intent-to-treat principle.
- Subject doesn't have any post-baseline assessment. The subject's data will be imputed as follows:

- For each subject at each scheduled assessment (Baseline to Week 32), a simple random sample with replacement is drawn from all the remaining subjects with the same treatment and with data at that scheduled assessment. The seed will be 7575.
 - Repeat the step 100 times.
 - The score of each subject is replaced by the mean score of the 100 samples drawn.
 - Subject is mis-stratified. The subject's data will be used as randomized based on the intent-to-treat principle.
4. Mean change from baseline in TSS7 based on MFSAF v4.0 at Week 32:

The weekly TSS7 will be calculated by averaging daily scores over a 7-day interval leading to Week 32 (inclusive). If at least 4 days are available, the score will be calculated.

The following events are defined as intercurrent events:

- Subject discontinues the study treatment prematurely during Part 1a for any reason. The subject's data will be used as provided. An MMRM approach will be employed to use all available visit data, including subjects with some missing visit data in order to arrive at an estimate of the mean treatment effect.
- Subject receives prohibited medications or prohibited changes in cytoreductive therapy between Baseline up to Week 32. The subject's data after the use of prohibited medications or prohibited changes in cytoreductive therapy will be considered missing. As a sensitivity analysis, the subject's data will be used as is.
- Subject doesn't comply with the treatment. The subject's data will be used as provided.
- Subject randomized but doesn't receive any treatment. The subject's data will be imputed as follows:
 - For each subject at each scheduled assessment (Baseline to Week 32), a simple random sample with replacement is drawn from all the remaining placebo subjects with data at that scheduled assessment. The seed will be 7707.
 - Repeat the step 100 times.
 - The score of each subject is replaced by the mean score of the 100 samples drawn.
- Subject receives the wrong treatment according to the randomization. The subject's data will be used as provided based on the intent-to-treat principle.
- Subject doesn't have any post-baseline assessment. The subject's data will be imputed as follows:
 - For each subject at each scheduled assessment (Baseline to Week 32), a simple random sample with replacement is drawn from all the remaining subjects with the same treatment and with data at that scheduled assessment. The seed will be 7575.

- Repeat the step 100 times.
- The score of each subject is replaced by the mean score of the 100 samples drawn.
- Subject is mis-stratified. The subject's data will be used as randomized based on the intent-to-treat principle.

Sensitivity to the model assumptions are evaluated in [Section 12.4](#) for the primary and key secondary endpoints, which may include multiple imputation and tipping point analyses.

For the other secondary endpoints:

In general, for the categorical secondary endpoints if a subject is missing a value for an assessment at a visit, he/she will be considered a non-responder for that endpoint at the specific time point. Missing data will not be imputed for calculation of mean changes and other quantitative analyses.

12.4. Sensitivity Analysis of the Primary and Key Secondary Endpoints

12.4.1. Per Protocol Analysis

The primary and the key secondary endpoints will be analyzed as outlined in [Section 12.2.1](#) and [Section 12.2.2](#) using the per protocol set (PPS) ([Section 10.4](#)).

12.4.2. Full Analysis Set Analysis

The primary and the key secondary endpoints will be analyzed as outlined in [Section 12.2.1](#) and [Section 12.2.2](#) using the full analysis set (FAS) ([Section 10.2](#)).

12.4.3. Tipping Point Analysis

Under Missing at Random (MAR) assumption, subjects who discontinue the treatment early are assumed to have the same data pattern as subjects who remain in the study. Patient-profile (spaghetti) plots will be generated for those who stay in the study, those who discontinue the treatment due to adverse event, those who drop out due to lack of efficacy, and those who drop out due to other reasons. Tipping point analysis, based on ITT, will be employed to assess the robustness of the results under the missing not at random (MNAR) assumption.

12.4.3.1. Categorical Endpoints: Primary Endpoint and Key Secondary Endpoint # 2

To assess the robustness of the results of the primary analysis with respect to missing hematocrit data, a multiple imputation with δ adjustment will be performed as follows:

- a. Use Fully Conditional Specification – predictive mean matching (for continuous variables) and logistic (for categorical variables) methods to impute missing HCT values at scheduled visits between Baseline and Week 32, inclusive. The model includes categorical variables: treatment group and randomization strata; continuous variables: baseline, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 28, and Week 32 hematocrit values. The missing data are filled in and 100 “complete” datasets are created. The seed that will be used in the SAS program is 770701.

- b. For each missing value, adjust the imputed value by adding a δ to the imputed value, where $\delta = 0$ (%).
Sample SAS codes for the multiple imputation steps (a) and (b) are provided in [Appendix 2](#).
- c. Use the observed and imputed hematocrit values to derive the response status (primary endpoint) of each subject (See [Sections 12.2.1](#) and [12.3](#) for the definition and derivation of the response status).
- d. Each of the 100 “complete” datasets will then be analyzed separately using the CMH test, stratified by ongoing PV therapy. The Mantel-Haenszel stratum-adjusted, common risk difference estimate and its associated standard error will be generated.
- e. The estimates obtained from the Mantel-Haenszel stratum-adjusted common risk difference and its standard error for each “complete” dataset are combined for inference purposes using Rubin’s rule ([Rubin 1987](#)).
- f. Repeat steps b, c, d, e with $\delta = 0.5, 1, 1.5, 2, 2.5, 3$ (%). The larger the value of δ that is required to reverse the conclusion of the primary analysis, the more robust the conclusion is. When $\delta = 0$ (no adjustment), this reduces to the standard multiple imputation under Missing at Random (MAR) assumption.

Similarly for the key secondary endpoint #2 (Proportion of subjects with all HCT <45% between Baseline and Week 32, inclusive), a multiple imputation procedure will be performed as above with step (c) is replaced with:

Use the observed and imputed hematocrit values to derive the response status (key secondary endpoint #2) of each subject. A subject will be considered a responder if all HCT <45% between Baseline and Week 32, inclusive. A single HCT $\geq 45\%$ is allowed.

12.4.3.2. Continuous Endpoint: Key Secondary Endpoints #1

To assess the robustness of the result of the analysis for key secondary endpoint #1 with respect to missing data, a multiple imputation with δ adjustment will be performed as follows:

- a. Use Fully Conditional Specification – predictive mean matching (for continuous variables) and logistic (for categorical variables) methods to impute the missing number of phlebotomies between Baseline and Week 32, inclusive. The model includes categorical variables: treatment group and randomization strata; continuous variables: prestudy phlebotomy rate (within 32 weeks prior to first dose date), duration of treatment (between Baseline and Week 32), and the number phlebotomy between Baseline and Week 32, inclusive. The missing data are filled in and 100 “complete” datasets are created. The seed that will be used in the SAS program is 770702.
- b. For each missing value, adjust the imputed value by a δ , where $\delta = 0$.
Sample SAS codes for the multiple imputation steps (a) and (b) are provided in [Appendix 2](#).
- c. Each of the 100 “complete” datasets will be analyzed separately using a model specified in [Section 12.2.2.1](#).
- d. The estimates and standard errors obtained from the model above are combined for inference purposes using Rubin’s rule ([Rubin 1987](#)).
- e. Repeat steps b, c, d with $\delta = 0.5, 1, 1.5, 2, 2.5, 3$. The larger the value of δ that is required to reverse the conclusion of the analysis, the more robust the conclusion is.

When $\delta = 0$ (no adjustment), this reduces to the standard multiple imputation under MAR assumption.

12.4.3.3. Continuous Endpoints: Key Secondary Endpoints #3, and #4

To assess the robustness of the results of the analyses for key secondary endpoints #3 and #4 with respect to missing data, a multiple imputation with δ adjustment will be performed as follows:

- a. Use Fully Conditional Specification – predictive mean matching (for continuous variables) and logistic (for categorical variables) methods to impute missing MFSAF TSS7 values at scheduled visits between Baseline and Week 32, inclusive. The model includes categorical variables: treatment group, randomization strata; continuous variables: baseline, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 28, and Week 32 MFSAF TSS7 values. The missing data are filled in and 100 complete datasets are created. The seed that will be used in the SAS program is 770703.
- b. For each missing value, adjust the imputed value by adding a δ to the imputed value, where $\delta = 0$ (point).

Sample SAS codes for the multiple imputation steps (a) and (b) for MFSAF TSS7 are provided in [Appendix 2](#). A similar code can be adapted for PROMIS T-Score (with less frequent visit assessments, i.e., remove Week 8, Week 16, and Week 24 because PROMIS are not administered at those visits) and MPN-SAF Problems with Concentration score.

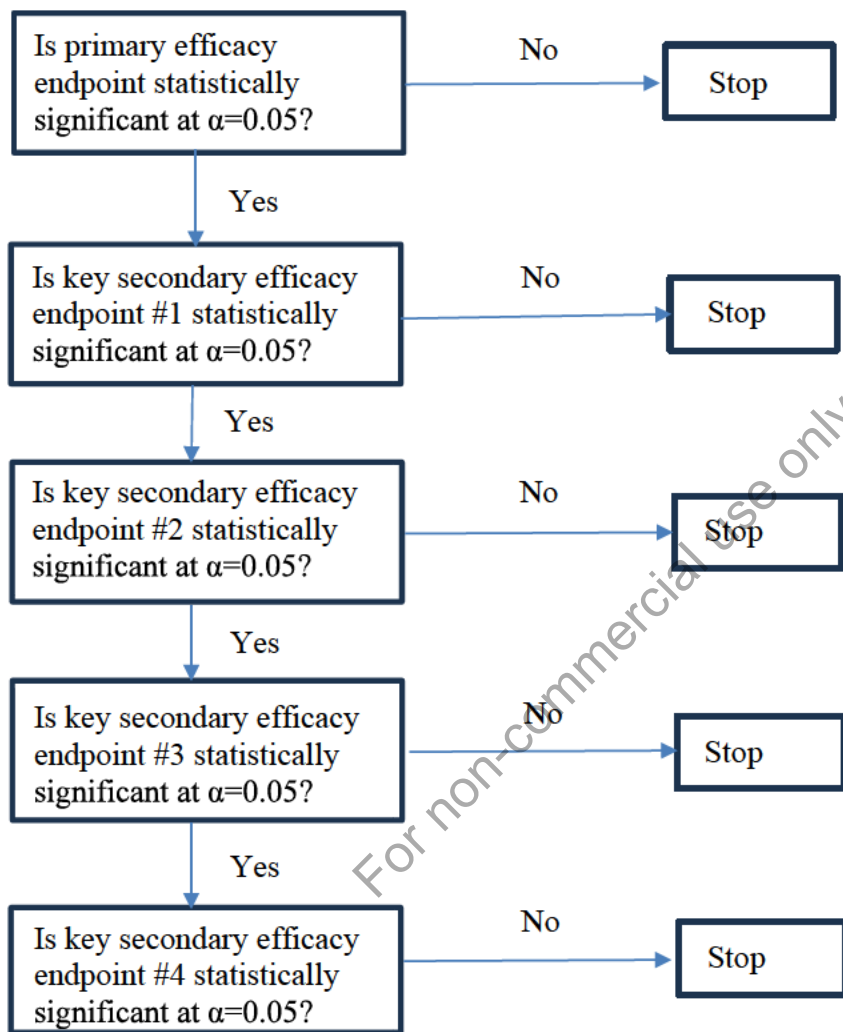
- c. Each of the 100 complete datasets will then be analyzed separately using an MMRM as specified in [Sections 12.2.2.3](#) and [12.2.2.4](#), respectively.
- d. The estimates and standard errors obtained from the MMRM for each complete dataset are combined for inference purposes using Rubin's rule ([Rubin 1987](#)).
- e. Repeat steps b, c, d with $\delta = 1, 2, 3, 4, 5$ (points) for MFSAF TSS7, PROMIS T-score. For MPN-SAF Problems with Concentrations, repeat steps b, c, d with $\delta = 1, 2, 3$. The larger the value of δ that is required to reverse the conclusion of the primary analysis, the more robust the conclusion is. When $\delta = 0$ (no adjustment), this reduces to the standard multiple imputation under MAR assumption.

12.4.4. Assessing Normality Assumptions of ANCOVA and MMRM Model

The normality assumptions of the ANCOVA and MMRM model will be examined graphically via histograms and normal probability plots of residuals. If the normality assumption is violated, then an analysis on rank may be considered.

12.5. Multiplicity Adjustment

The fixed-sequence method will be employed to address multiple endpoint-related multiplicity. The endpoints are ordered in a predefined sequence and the family-wise error rate will be controlled at 0.05 level of significance for the primary endpoint and the 4 key secondary endpoints. The testing sequence is as follows:



12.6. Examination of Subgroups

Subgroup analysis currently planned for this study includes the following. This list may be expanded as more subgroups are identified for further analyses.

- Sex (Male/Female)
- Race (White/Others)
- Age (<65 , ≥ 65 years)
- Thromboembolic risk category (High [age ≥ 60 years or prior thromboembolic event], Low [age < 60 years and no prior thromboembolic event])
- Region (North America, Europe, Others)
- Prior cytoreductive therapy (Y/N)
- Ongoing cytoreductive therapy (Y/N)
- Duration of PV (≤ 3 years, > 3 years)
- Phlebotomy rate prior to enrollment (annualized) (≤ 6 /year vs > 6 /year)
- BMI (< 30 kg/m², ≥ 30 kg/m²)

For non-commercial use only

13. CLINICAL PHARMACOLOGY ANALYSIS

13.1. Pharmacokinetic (PK) Analysis

All PK analyses will be conducted using the PKAS.

Sparse blood sampling will be conducted for plasma concentrations. Given the sparse nature of sampling, noncompartmental pharmacokinetic analysis will not be conducted.

Individual rusfertide concentration data will be listed by subject, nominal time point, and summarized by time point with descriptive statistics (sample size, arithmetic mean, SD, median, minimum, maximum, geometric mean, and arithmetic coefficient of variation). Concentration data may be further segmented by dose if appropriate. All plasma concentration values below the limit of quantification (BLQ) will be set to zero, except for geometric mean, when calculating summary statistics. For geometric mean, plasma concentration values below the BLQ will be set to the lower limit of quantification. Individual and mean rusfertide concentration-time profiles may be presented graphically.

PK blood sample will be collected at the following time points:

- Predose: Weeks 0, 8, 16, 24, 32, 36, 40, 52, 60, 72, 84, 96, 108, 120, 132, 144, 156/EOT, EOS
- Postdose: Weeks 16, 52, 108

Trough concentration at Weeks 8, 16, 24, 32, 36, 40, 52, 56, 72, 84, 96, 120, 132, 144, 156 (end of treatment) predose, at 4 weeks following last dose and at early termination. Post-dose concentrations at Weeks 16, 52, and 108 will also be summarized, if available.

13.2. Pharmacodynamic (PD) Analysis

All PD analyses will be conducted on the PDAS population.

TSAT, serum iron, serum ferritin, and MCV will be summarized in tables using descriptive statistics. Change from baseline will also be summarized for each available time point.

Individual pharmacodynamic data will be listed.

13.3. Exploratory Biomarkers of Interest

Exploratory biomarkers to be assessed will include, but are not limited to, the following:

- JAK2 mutation burden

14. SAFETY ASSESSMENTS

The Safety Analysis Set will be used for analysis of all safety data.

Quantitative safety data will be summarized using descriptive statistics (n, arithmetic mean, SD, median, minimum, and maximum). Categorical safety data will be summarized by frequencies and percentages.

Adverse events (AEs) will be coded using Medical Dictionary for Regulatory Activities (MedDRA®), Concomitant medications will be coded using World Health Organization (WHO) drug dictionary.

Laboratory evaluations, vital signs assessments and ECG parameters will be summarized by treatment group and protocol specified collection time point. A summary of change-from-baseline at each protocol specified time point will also be presented.

No statistical tests will be performed for any of the safety assessments. All subject level safety data will be listed.

14.1. Extent of Exposure and Compliance

Frequency of the various dosing regimen and dosing at End of Treatment will be summarized. Dosing regimen will be per eCRF Dose Regimen form.

The duration of exposure and weekly dose will be determined. In general, for the overall exposure the following equation will be used.

- Duration of exposure [weeks] = (Date/time of last dose–Date/time of first dose+1)/7 and round the result to the lower integer. Additionally, 6 days will be added to duration of exposure if the regimen is weekly, 3 days if the dosing regimen is twice a week, 13 days if the dosing regimen is once every 2 weeks.
- Dose (in mg)/Week =
 - dose administered if “weekly” dosing regimen.
 - (dose on Day 1 + dose on Day 3 or 4) if “two times a week” dosing regimen.
 - dose/2 if subject is “every two weeks” dosing regimen.

Mean weekly dose over the study period will be calculated using descriptive statistics and presented in a table for the Safety Analysis Set.

Drug compliance will be calculated per the Drug Accountability Form. Compliance will be calculated as 1 minus the ratio of number of unused syringes returned over the number of syringes dispensed. For subjects who were injected with diluents only, compliance will be calculated removing the period of subjects taking the diluents only. If there are discrepancies between the Dosing Diary and the Drug Accountability, data from the Drug Accountability will be used for the compliance calculation purposes.

14.2. Adverse Events

14.2.1. Adverse Event Dictionary

AEs will be coded using MedDRA. System organ class, high-level group term (HLGT), high-level term (HLT), preferred term (PT), and lowest-level term (LLT) will be included in the AE dataset.

14.2.2. Adverse Event Severity

Adverse events will be graded by the investigators using the CTCAE version 5.0, based on the eCRF question “CTCAE version 5.0 grade.” Categories include Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe), Grade 4 (life-threatening), Grade 5 (fatal). AEs with missing severity will be queried and if remain missing, will be categorized as “missing” in tabular summary.

14.2.3. Relationship of Adverse Event to Study Drug

Relationship to study drug will be based on the eCRF question “Relationship to Study Treatment” and will be dichotomized as “Related” if the answer to the question above is “Related” or “Possibly Related” and as “Not related,” if otherwise. Adverse events with a missing relationship will be queried and if unresolved, will be considered related to study drug.

14.2.4. Serious Adverse Event

A serious adverse event (SAE) is any untoward medical occurrence that:

- Results in death
- Is life-threatening
- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in persistent or significant disability/incapacity, or
- Is a congenital anomaly/birth defect
- Important medical events that are not one of the above may be considered SAEs by the investigator when, based upon appropriate medical judgement, they are considered to be clinically significant and may jeopardize the subject, or when medical or surgical intervention may be required to present one of the outcomes listed above.

SAEs will be identified in the eCRF.

14.2.5. Treatment-Emergent Adverse Events

For the study duration:

Treatment-emergent AEs (TEAEs) are defined as those AEs that begin on or after first study drug administration at the start of the study up to 10 days after last dose of study drug. For cancer events, all events that begin on or after the first study drug administration will be considered treatment-emergent.

In particular,

For Part 1a:

Treatment-emergent AEs (TEAEs) are defined as those AEs that begin on or after first study drug of Part 1a administration at the start of the study up to 10 days after last dose of part 1a study drug or up to the first dose of Part 1b study drug, whichever is first. For ‘skin lesion possibly malignant’ and confirmed to be skin cancers that are discovered during the Week 32 dermat exam (Week 32 visit + 10 days), the event will be attributed to Part 1a.

For Part 1b:

Treatment-emergent AEs (TEAEs) are defined as those AEs that begin on or after first study drug of Part 1b administration up to 10 days after last dose of Part 1b study drug or up to the first dose of Part 2 study drug, whichever is first.

For Part 2:

Treatment-emergent AEs (TEAEs) are defined as those AEs that begin on or after first study drug of Part 2 administration up to 10 days after last dose of Part 2 study drug.

If it cannot be determined whether the AE is treatment emergent due to a partial onset date, then the date will be imputed based on rules in [Appendix 1](#).

Only TEAEs are tabulated. AE listings will present all AEs (treatment-emergent and non-treatment-emergent).

14.2.6. Summary of Adverse Events

Summary of AEs will be provided for the Safety Analysis Set by treatment group and by Part 1a, 1b, and 2 and all rufertide. The number and percentage of subjects who experienced the following will be provided by SOC, PT:

- TEAEs
- Serious TEAEs
- Treatment-related TEAEs
- Treatment-related serious TEAEs
- TEAEs by severity
- Serious TEAEs by severity
- Treatment-related TEAEs by severity
- Treatment-related Serious TEAEs by severity
- TEAEs leading to premature discontinuation from the treatment
- Serious TEAEs leading to premature discontinuation from the treatment

A brief overall summary of AEs described above will be provided for TEAEs, treatment-related TEAEs, TEAEs leading to discontinuation, treatment-emergent SAEs, treatment-related treatment-emergent SAEs, treatment-emergent SAEs leading to discontinuation.

Multiple events will be counted only once per subject in each summary. Summaries displayed by SOC and PT will be ordered in alphabetical order of SOC and descending incidence of PT within each SOC.

In addition to the above summary tables, all TEAEs and treatment-related TEAEs will be summarized by PT only and will be ordered in descending incidence of PT to identify the frequently occurred TEAE by PT.

Separate tables will be generated for TEAEs which meet the following criteria:

- AESIs, including the following:
 - Cancer events are searched using the FDA Medical Queries (FMQ) – narrow terms for malignancy -excluding preferred term myeloproliferative neoplasm.

In addition, incidence of cancer events will be calculated per 1000 person-year follow-up for the two treatment groups during Part 1a, Part 1a + Part 1b. Incidence of cancer events will be calculated similarly for the whole study.

For subjects with at least one event of interest, the formula is:

$$\text{Person Year} = \frac{\text{Onset date of the first event} - \text{date of start of treatment} + 1}{365.25}$$

For subjects without any event, the formula is:

$$\text{Person Year} = \frac{\text{Duration of exposure}}{365.25}$$

The total at risk time is the sum of all subjects at risk exposure time. The incidence per 1000 person-year follow-up is calculated as the ratio of number of subjects experiencing the event over the total at risk time for the event in the population, the formula is:

$$\frac{\text{Number of subjects with events}}{\text{Total Person Year}} \times 1000$$

95% confidence interval will be computed for each treatment group.

- TEAEs marked as Injection Site Reactions (ISRs). The TEAE will be marked as an ISR if the site identifies the TEAE as an injection site reaction by checking the appropriate box on the eCRF (“Check if the event is related to an injection site reaction”).
- TEAEs identified as PV disease progression include the following:
 - Myelofibrosis
 - Leukemia
 - Myelodysplastic syndrome
- TEAEs identified as PV complications. PV complications include thromboembolic events. These events will be searched using the FMQ-narrow terms for Thrombosis.

Tabulation of subjects with prior history of thromboembolic events in medical history and have reported a thromboembolic event during the study will be provided.

Incidence of thromboembolic events will be calculated per 1000 person-year follow-up for the two treatment groups during Part 1a, Part 1a+Part 1b, and Part 1a+ Part 1b + Part 2.

- TEAEs by prior or concurrent exposure to cytoreductive:
 - Prior (but not concurrent) exposure to cytoreductive
 - Prior and concurrent exposure to cytoreductive
 - Concurrent (but not prior) exposure to cytoreductive
 - No prior nor concurrent exposure to cytoreductive

Listings will be provided for the following:

- All AEs
- All SAEs
- Treatment-related AEs
- Treatment-related SAEs
- AEs of Grade 3 or higher
- AEs leading to discontinuation of treatment. For AEs with an onset prior to an epoch and the AE leads to discontinuation of treatment during an epoch, they will be listed in the epoch that the subject discontinues the treatment.
- AESIs
- ISRs
- AEs identified as PV Disease Progression
- AEs identified as PV complications
- Grade 3 or above bleeding AEs (HLGT = Vascular Haemorrhagic Disorders) associated simultaneously (within +/- 7 days of the onset of the AE) with platelets $>1500 \times 10^9/L$
- Subjects with anemia AE or HGB ≤ 10.5 g/L

14.3. Clinical Laboratory Tests

A complete list of clinical laboratory tests is provided in Appendix 3 of the protocol. Laboratory data will be analyzed and summarized using both quantitative and qualitative methods by scheduled time points. Summaries of laboratory data will be provided for the Safety Analysis Set. The analysis will be based on values reported in the international system of unit (SI). If the reported values are not in SI, they will be converted to SI before analyses are performed. When values are below or above the detection limit, they will be listed as such and will be imputed by the detection limit value itself for the purpose of calculating summary statistics. Example: if the database contains a value of “<0.04”, then for the descriptive statistics the value of the detection limit (0.04) will be used. A value such as “>1000” will be imputed by “1000” for summary tables. All lab results that are fractions of 1 will be converted to percent. When more than one

reference range is available due to different local labs used, the most stringent reference range will be listed for the derived values.

Descriptive statistics will be provided by subject group for each laboratory tests as follows:

- Baseline values
- Values at each postbaseline time point
- Change from baseline at each postbaseline time point (for continuous parameters)

Shift tables (3×3 tables) will be provided to assess changes in laboratory values from baseline to each scheduled visit.

Reference ranges and Clinically Significant (CS) information will be collected. Subject listings of all safety laboratory data collected during the study will be presented. Listings will indicate those values outside the reference ranges and if clinically significant.

14.3.1. Hematology

The following hematology tests will be summarized using descriptive statistics in a table using the Safety Analysis Set:

- White blood cell (WBC) parameters: WBC count, basophils, eosinophils, lymphocytes, monocytes, and neutrophils.
- Red blood cell (RBC) parameters: Hemoglobin, hematocrit, and RBC count.
- Platelet count.
- Mean corpuscular volume (MCV), mean corpuscular hemoglobin content (MCHC), mean corpuscular hemoglobin (MCH)

Note that for WBC differentials (basophils, eosinophils, lymphocytes, monocytes, neutrophils), SI units are considered counts, not % values, only the count result will be summarized. The % values will be listed.

The following analyses will be conducted for the entire duration of study drug exposure and will be summarized by Part 1a, 1b, 2, and overall:

- Proportion of subjects at the last on-treatment visit meeting all the criteria listed below:
 - WBC $<10 \times 10^9/L$
 - Platelets $<400 \times 10^9/L$
 - Hematocrit $<45\%$
- Proportion of subjects meeting the following at any visit:
 - Anemia
 - Grade 1 – Hemoglobin (HGB) between 100 g/L and $<$ lower limit of normal
 - Grade 2 – HGB between 80 g/L and <100 g/L
 - Grade 3 – HGB <80 g/L

- Platelet count decreased
 - Grade 1 – Platelets between $75 \times 10^9/L$ and $< \text{lower limit of normal}$
 - Grade 2 – Platelets between $50 \times 10^9/L$ and $< 75 \times 10^9/L$
 - Grade 3 – Platelets between $25 \times 10^9/L$ and $< 50 \times 10^9/L$
 - Grade 4 – Platelets $< 25 \times 10^9/L$
 - White blood cell decreased
 - Grade 1 – WBC between $3 \times 10^9/L$ and $< \text{lower limit of normal}$
 - Grade 2 – WBC between $2 \times 10^9/L$ and $< 3 \times 10^9/L$
 - Grade 3 – WBC between $1 \times 10^9/L$ and $< 2 \times 10^9/L$
 - Grade 4 – WBC $< 1 \times 10^9/L$
 - Neutrophil count decreased
 - Grade 1 – Neutrophils between $1.5 \times 10^9/L$ and $< 2 \times 10^9/L$
 - Grade 2 – Neutrophils between $1 \times 10^9/L$ and $< 1.5 \times 10^9/L$
 - Grade 3 – Neutrophils between $0.5 \times 10^9/L$ and $< 1 \times 10^9/L$
 - Grade 4 – Neutrophils $< 0.5 \times 10^9/L$
 - Proportion of subjects with:
 - Continuous (all assessments) platelet counts $> 1000 \times 10^9/L$ for at least 8 weeks
 - Continuous (all assessments) platelet counts $> 1500 \times 10^9/L$ for at least 8 weeks
 - Continuous (all assessments) WBC $> 15 \times 10^9/L$ for at least 8 weeks
 - Continuous (all assessments) WBC $> 25 \times 10^9/L$ for at least 8 weeks
- For this analysis, if the span of 8 weeks crosses from Part 1a to Part 1b or from Part 1b to Part 2, then the proportion will be counted toward the beginning Part.

14.3.2. Serum Chemistry

The following chemistry tests will be summarized using descriptive statistics in a table for the Safety Analysis Set:

- Hepatobiliary parameters: Alanine transaminase (ALT), aspartate transaminase, alkaline phosphatase (ALP), total bilirubin, indirect bilirubin, and direct bilirubin.
- Renal function tests: Blood urea nitrogen, serum creatinine, estimated glomerular filtration rate (eGFR).
- General serum chemistry: Albumin, bicarbonate, glucose, sodium, potassium, chloride, phosphorous, total protein, uric acid, calcium, lactate dehydrogenase (LDH).

14.3.3. Urinalysis

The following urinalysis tests will be listed using the Safety Analysis Set:

- pH, specific gravity, blood, glucose, ketones, leukocyte esterase, nitrite, protein, bilirubin, urobilinogen. Microscopic examination results will be listed if performed.

14.3.4. Additional Lab Tests

The following additional tests will be listed using the Safety Analysis Set:

- Hepatitis B (HBsAG, HBc) and C, HIV, pregnancy test when data are available.
- Reticulocytes, red cell distribution width (RDW), and inflammation markers when data are available.
- Coagulation: International normalized ration (INR), prothrombin time (PT), partial thromboplastin time (PTT).

14.4. Prior and Concomitant Medications

Prior and concomitant medication verbatim terms in the eCRFs will be mapped to Anatomical/Therapeutic/Chemical (ATC) class and Preferred Names using the WHODrug.

- Previous medications are defined with a stop date before the date of the start of study medication.
- Concomitant medication are medications taken during or after study drug administration (e.g., start date of medication is on or after first dose date) or medications started before study drug administration and ended after study drug administration or ongoing.

Prior and concomitant medications will be summarized for the overall study by WHO ATC class level 4, or the next level available if ATC level 4 is missing, and preferred name using the Safety Analysis Set. These summaries will present the number and percentage of subjects using each medication. Subjects may have more than 1 medication per ATC class and preferred name. At each level of subject summary, a subject is counted once if he/she reported 1 or more medications at that level. Each summary will be ordered in alphabetical order of ATC class and by descending order of incidence of preferred name within each ATC class. For drugs with the same frequency, sorting will be done alphabetically. All prior and concomitant medications will also be listed.

If only partial start and end dates are available, then the date will be imputed using rules described in [Appendix 1](#).

14.5. Vital Signs

Vital signs will be summarized in a table using descriptive statistics at baseline and at each postbaseline time point when data are collected using the Safety Analysis Set. Summaries will include:

- Baseline value
- Values at each postbaseline time point
- Change from baseline at each postbaseline time point.

Parameters to include in vital signs descriptive statistics tables are:

- weight (kg),

- BMI (kg/m²)
- temperature (°C),
- heart rate (beats/min)
- respiration rate (breaths/min),
- blood pressure: systolic blood pressure (SBP) (mmHg), diastolic blood pressure (DBP) (mmHg)

In addition, the number and percentage of subjects with any out-of-range postbaseline vital sign value will be summarized by scheduled visit as specified below:

- Change in weight from baseline:
 - $\geq 5\%$ increase
 - $\geq 5\%$ decrease
- Temperature:
 - $> 38.3^{\circ}\text{C}$
 - $\geq 1.1^{\circ}\text{C}$ increase from baseline
- Heart Rate:
 - < 50 bpm
 - > 120 bpm
- Respiration Rate:
 - < 8 breaths/min
 - > 40 breaths/min
- Blood Pressure:
 - SBP > 150 mmHg or DBP > 100 mmHg
 - SBP > 180 mmHg or DBP > 110 mmHg

14.6. 12-Lead ECG

12-Lead ECGs will be performed and interpreted. Overall interpretation results for ECG will be summarized using shift tables (Normal, Abnormal Not Clinically Significant, Abnormal Clinically Significant) comparing baseline to follow-up by each scheduled visit using the Safety Analysis Set. Summaries will be presented by treatment group.

Descriptive statistics at baseline and at each postbaseline time point as well as changes from baseline will be summarized for ECG parameters: ventricular rate (bpm), PR Interval (msec), QRS (msec), QT interval (msec), RR interval (msec), corrected QT interval based on Fridericia's formula (QTcF [msec]) and Rhythm Assessment.

The summary of number and percentage of subjects for rhythm assessment (Normal Sinus Rhythm VR 60-100 bpm, Sinus bradycardia VR less than 60 bpm, Sinus tachycardia VR greater than 100 bpm, and other) will also be included. In addition, a categorical summary of abnormal

corrected QT interval (QTc) values with the highest postbaseline result or highest change from baseline value will be presented for the categories below:

- The number and percentage of subject with highest postbaseline QTcF values of below categories:
 - ≤ 450 msec
 - >450 msec to ≤ 480 msec
 - >480 msec to ≤ 500 msec
 - >500 msec
- The number and percentage of subject with highest change from baseline values in QTcF of below categories:
 - >30 msec to ≤ 60 msec
 - >60 msec

14.6.1. Change in Spleen Size from Screening

The spleen size and change from screening in spleen size by radiology will be summarized and listed.

14.7. Immunogenicity

The proportion of subjects developing anti-drug antibodies (ADA) will be summarized by visit and overall in a table for the Safety Analysis Set. All ADA data including titers when available will be listed.

15. REFERENCES

Cella D, Riley W, Stone A, et al. Initial Adult Health Item Banks and First Wave Testing for the Patient-Reported Outcomes Measurement Information System (PROMIS™) Network: 2005-2008. *J Clin Epidemiol*. 2010 Nov;63(11): 1179-1194.

Emanuel RM, Dueck AC, Geyer HL, et al. Myeloproliferative neoplasm (MPN) symptom assessment form total symptom score: prospective international assessment of an abbreviated symptom burden scoring system among patients with MPNs [published correction appears in *J Clin Oncol*. 2012 Dec 20;30(36):4590. Ferarri, Maria L [corrected to Ferrari, Maria L]]. *J Clin Oncol*. 2012;30(33):4098-4103. doi:10.1200/JCO.2012.42.3863

Food and Drug Administration. FDA Medical Queries. Silver Spring, Maryland, 2023 (<https://www.fda.gov/media/164639/download>)

Gwaltney C, Paty J, Kwitkowski VE, et al. Development of a harmonized patient-reported outcome questionnaire to assess myelofibrosis symptoms in clinical trials. *Leuk Res*. 2017;59:26-31. doi:10.1016/j.leukres.2017.05.012

Ivanova A, Qaqish BF, Schell MJ. Continuous toxicity monitoring in phase II trials in oncology. *Biometrics*. 2005;61(2):540-545. doi:10.1111/j.1541-0420.2005.00311.x

Khanal N, Giri S, Upadhyay S, Shostrom VK, Pathak R, Bhatt VR. Risk of second primary malignancies and survival of adult patients with polycythemia vera: A United States population-based retrospective study. *Leuk Lymphoma*. 2016;57(1):129-133. doi:10.3109/10428194.2015.1071492

Masarova L, Cherry M, Newberry KJ, et al. Secondary solid tumors and lymphoma in patients with essential thrombocythemia and polycythemia vera - single center experience. *Leuk Lymphoma*. 2016;57(1):237-239. doi:10.3109/10428194.2015.1041392

Quality of Life Unit, EORTC Data Center. EORTC QLQ-C30 Scoring Manual. Brussels, Belgium, 2001 (<https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf>).

Rubin DB. Multiple Imputation for Nonresponse in Surveys. San Francisco, CA: John Wiley & Sons, Inc; 1987. doi: 10.1002/9780470316696

Vannucchi AM, Kiladjian JJ, Griesshammer M, et al. Ruxolitinib versus standard therapy for the treatment of polycythemia vera. *N Engl J Med*. 2015;372(5):426-435. doi:10.1056/NEJMoa1409002

APPENDIX 1. MISSING OR INCOMPLETE DATE IMPUTATION

Start Dates

- If the day portion of the start date is missing, then the start date will be estimated to be equal to the date of first dose of study drug, provided the start month and year are the same as the first dose of study drug and the stop date is either after the first dose of study drug or completely missing. Otherwise, the missing day portion will be estimated as “01.”
- If both the day and month portions of the start date are missing, then the start date will be estimated to be equal to the date of first dose of study drug, provided the start year is the same as the first dose of study drug and the stop date is either after the first dose of study drug or completely missing. Otherwise, the event will be assumed to start on the first day of the given year (e.g., unk-unk-2022 is estimated as 01-JAN-2022).
- If the start date is completely missing and the stop date is either after the dose of study drug or completely missing, the start date will be estimated to be the day of study drug dosing. Otherwise, the start date will be estimated to be the first day of the same year as the stop date.
- If the first dose date is not available, then the randomization date will be used instead of the first dose date.
- If the start date and stop date are completely missing and the ongoing status is No, then it is assumed that the event occurs and resolves before the first dose of study drug.

Stop Dates

- If only the day of resolution is unknown, the day will be assumed to be the last of the month (e.g., unk-JAN-2022 will be treated as 31-JAN-2022).
- If both the day and month of resolution are unknown, the event will be assumed to have resolved on the last day of the year (e.g., unk-unk-2022 will be treated as 31-DEC-2022) or the last known date on the study, whichever is earlier.
- If the stop date is completely missing or the event is continuing, the event will be assumed to be after first dose of study drug and will be imputed using the last known date on the study.

APPENDIX 2. SAMPLE SAS CODES FOR MULTIPLE IMPUTATION

```
** Multiple imputation macro;
%macro mi(indata = ,
          var0 = ,
          var4 = ,
          var8 = ,
          var12 = ,
          var16 = ,
          var20 = ,
          var24 = ,
          var28 = ,
          var32 = ,
          catvar = , seed = , delta = , nimpute = );

%if &var8 ^= "" %then %do;
*Imputation step;
proc mi data = &indata seed = &seed nimpute = &nimpute out = out1_&indata ;
  class &catvar;
  fcs logistic (&catvar/link = glogit) ;
  fcs regpmm (&var0 &var4 &var8 &var12 &var16 &var20 &var24 &var28 &var32);
  var &catvar &var0 &var4 &var8 &var12 &var16 &var20 &var24 &var28 &var32;
run;

*rename the variables in original datasets;
data &indata;
  set &indata;
  rename &var0 = _&var0
         &var4 = _&var4
         &var8 = _&var8
         &var12 = _&var12
         &var16 = _&var16
         &var20 = _&var20
         &var24 = _&var24
         &var28 = _&var28
         &var32 = _&var32;
run;

*Adjust imputed values by delta;
proc sort data = &indata; by usubjid; run;
proc sort data = out1_&indata; by usubjid; run;

data out2_&indata;
  merge &indata out1_&indata ;
  by usubjid;
  if _&var0 = . then &var0 = &var0 + &delta;
  if _&var4 = . then &var4 = &var4 + &delta;
  if _&var8 = . then &var8 = &var8 + &delta;
  if _&var12 = . then &var12 = &var12 + &delta;
  if _&var16 = . then &var16 = &var16 + &delta;
  if _&var20 = . then &var20 = &var20 + &delta;
  if _&var24 = . then &var24 = &var24 + &delta;
  if _&var28 = . then &var28 = &var28 + &delta;
  if _&var32 = . then &var32 = &var32 + &delta;
run;
%end;
```

```

*For PROMIS;
%if &var8 = "" %then %do;
*Imputation step;
proc mi data = &indata seed = &seed nimpute = &nimpute out = out1_&indata ;
  class &catvar;
  fcs logistic (&catvar/link = glogit) ;
  fcs regpmm (&var0 &var4 &var12 &var20 &var28 &var32);
  var &catvar &var0 &var4 &var12 &var20 &var28 &var32;
run;

*rename the variables in original datasets;
data &indata;
  set &indata;
  rename &var0 = _&var0
         &var4 = _&var4
         &var12 = _&var12
         &var20 = _&var20
         &var28 = _&var28
         &var32 = _&var32;
run;

*Adjust imputed values by delta;
proc sort data = &indata; by usubjid; run;
proc sort data = out1_&indata; by usubjid; run;

data out2_&indata;
  merge &indata out1_&indata ;
  by usubjid;
  if _&var0 = . then &var0 = &var0 + &delta;
  if _&var4 = . then &var4 = &var4 + &delta;
  if _&var12 = . then &var12 = &var12 + &delta;
  if _&var20 = . then &var20 = &var20 + &delta;
  if _&var28 = . then &var28 = &var28 + &delta;
  if _&var32 = . then &var32 = &var32 + &delta;
run;
%end;
%mend;

```

Hematocrit Endpoints

```

%mi(indata = hct,
    var0 = hct0,
    var4 = hct4,
    var8 = hct8,
    var12 = hct12,
    var16 = hct16,
    var20 = hct20,
    var24 = hct24,
    var28 = hct28,
    var32 = hct32,
    catvar = trtpn stratirt,
    seed = 770701, delta = 1, nimpute = 100);
*where hct0 = hematocrit at Baseline
      Hct4 = hematocrit at Week 4
      Hct8 = hematocrit at Week 8
      Hct12 = hematocrit at Week 12

```

```
Hct16 = hematocrit at Week 16
Hct20 = hematocrit at Week 20
Hct24 = hematocrit at Week 24
Hct28 = hematocrit at Week 28
Hct32 = hematocrit at Week 32
delta is the shift parameter and ranges from 0.5, 1, 1.5, 2, 2.5, 3
trtpn = treatment group
stratirt = randomization strata;
```

MFSAF Endpoint

```
%mi(indata = mfsaf,
    var0 = mfstot0,
    var4 = mfstot4,
    var8 = mfstot8,
    var12 = mfstot12,
    var16 = mfstot16,
    var20 = mfstot20,
    var24 = mfstot24,
    var28 = mfstot28,
    var32 = mfstot32,
    catvar = trtpn stratirt,
    seed = 770703, delta = 1, nimpute = 100);
*where mfstot0 = MFSAF TSS7 at Baseline
Mfstot4 = MFSAF TSS7 at Week 4
Mfstot8 = MFSAF TSS7 at Week 8
Mfstot12 = MFSAF TSS7 at Week 12
Mfstot16 = MFSAF TSS7 at Week 16
Mfstot20 = MFSAF TSS7 at Week 20
Mfstot24 = MFSAF TSS7 at Week 24
Mfstot28 = MFSAF TSS7 at Week 28
Mfstot32 = MFSAF TSS7 at Week 32
delta is the shift parameter and ranges from 1, 2, 3, 4, 5
trtpn = treatment group
stratirt = randomization strata;
```

PROMIS Endpoint

```
%mi(indata = prmtscr,
    var0 = prmtscr0,
    var4 = prmtscr4,
    var8 = "",
    var12 = prmtscr12,
    var16 = "",
    var20 = prmtscr20,
    var24 = "",
    var28 = prmtscr28,
    var32 = prmtscr32,
    catvar = trtpn stratirt,
    seed = 770703, delta = 1, nimpute = 100);
*where prmtscr0 = PROMIS T-score at Baseline
prmtscr4 = PROMIS T-score at Week 4
```

```
prmtscr12 = PROMIS T-score at Week 12
prmtscr20 = PROMIS T-score at Week 20
prmtscr28 = PROMIS T-score at Week 28
prmtscr32 = PROMIS T-score at Week 32
delta is the shift parameter and ranges from 1, 2, 3, 4, 5
trtpn = treatment group
stratirt = randomization strata;
```

MPN-SAF Problems with Concentration Endpoint

```
%mi(indata = mpn105,
    var0 = mpn1050,
    var4 = mpn1054,
    var8 = mpn1058,
    var12 = mpn1012,
    var16 = mpn1016,
    var20 = mpn1020,
    var24 = mpn1024,
    var28 = mpn1028,
    var32 = mpn1032,
    catvar = trtpn stratirt,
    seed = 770703, delta = 1, nimpute = 100);
*where mpn1050 = MPN-SAF Problems with Concentration at Baseline
    mpn1054 = MPN-SAF Problems with Concentration at Week 4
    mpn1058 = MPN-SAF Problems with Concentration at Week 8
    mpn10512 = MPN-SAF Problems with Concentration at Week 12
    mpn10516 = MPN-SAF Problems with Concentration at Week 16
    mpn10520 = MPN-SAF Problems with Concentration at Week 20
    mpn10524 = MPN-SAF Problems with Concentration at Week 24
    mpn10528 = MPN-SAF Problems with Concentration at Week 28
    mpn10532 = MPN-SAF Problems with Concentration at Week 32
    delta is the shift parameter and ranges from 1, 2, 3
    trtpn = treatment group
    stratirt = randomization strata;
```

Phlebotomy Endpoints

```
proc mi data = phl seed = 770702 nimpute = 100 out = phl1;
class TRTPN STRATIRT;
fcs logistic (TRTPN STRATIRT/link = glogit) ;
fcs regpmm (RPHL32 TRTDUR NTRTPL32_);
var TRTPN STRATIRT RPHL32 TRTDUR NTRTPL32_;
run;

*Rename the variables in original datasets;
data phl;
    set phl;
    rename NTRTPL32_ = NTRTPL32_tmp;
run;

proc sort data = phl; by usubjid; run;
proc sort data = phl1; by usubjid; run;

*Adjust imputed values by delta;
%let delta = 1;
```

```
data phl2;  
  merge phl phl1;  
  by usubjid;  
  if ntrtpl32_tmp = . then ntrtpl32_ = ntrtpl32_ + &delta;  
run;  
  *where TRTPN is the treatment group;  
  *STRATAIRT: Randomization strata;  
  *RPHL32: Phlebotomy rate 32 weeks prior to first dose;  
  *TRTDUR: Duration of treatment (between Baseline and Week 32);  
  *NTRTPL32_: Number of phlebotomy between Baseline and Week 32;  
  *delta is the shift parameter and ranges from 0 to 3;
```

For non-commercial use only

APPENDIX 3. DEFINING A SYMPTOMATIC SUBGROUP OF PATIENTS FOR THE PV-FOCUSED PRO SECONDARY ENDPOINTS

Protagonist has defined three PV-focused PRO secondary endpoints using the PROMIS Fatigue Short Form 8a total T-score, MFSAF TSS4, and the MPN-SAF problems with concentration symptom score. All three endpoints specify testing in a subset of patients who are symptomatic at baseline. This is because patients with PV (the target population for rusfertide) are characteristically less symptomatic than patients with more progressive MPNs (e.g., myelofibrosis), and as there were no inclusion criteria requiring patients to be symptomatic in Study PTG-300-11, approximately one-third of patients enrolled reported no symptoms. A further 43% of patients had only mild symptoms based on the PGI-S. Including non-symptomatic patients in the endpoints may not allow for meaningful differences between rusfertide and placebo to be well understood. As such, Protagonist undertook a statistical exploration of blinded baseline data from Study PTG-300-11 to determine an empirical definition of “symptomatic at baseline” which could be used to define a subgroup of patients for these endpoints.

Specifically, Protagonist evaluated various definitions of “symptomatic at baseline” for the PROMIS Fatigue Short Form 8a total score, MFSAF TSS4, and MPN-SAF problems with concentration symptom score using a blinded database cut from the PTG-300-11 Phase 3 study dated 23 July 2024. This dataset included randomized and dosed patients with baseline data for the PROMIS Fatigue Short Form 8a, MPN-SAF, and MFSAF. Six approaches were taken to potentially define patients who were “symptomatic at baseline”:

1. Those with a PGI-S baseline score of moderate or severe
2. Those with a baseline score on the outcome measure (PROMIS Fatigue Short Form 8a total T-score, MFSAF TSS4, and problems with concentration symptom score from the MPN-SAF) corresponding to moderate or severe using Receiver Operating Characteristic (ROC) derived cut-offs where the following approach was taken:
 - A series of logistic regressions using the outcome measure score as the independent variable and a binary variable (“none” or “mild” vs “moderate” or “severe”) as the dependent variable. The optimal cut-off values for different levels of symptom severity were determined by maximizing Youden’s J statistic (the sum of sensitivity and specificity minus one) when the area under the curve (AUC) exceeded 0.70.
3. Those with a baseline score on the outcome measure corresponding to moderate or severe using descriptive analysis where the following approach was taken:
 - The distribution of the outcome measure score was categorized at baseline using the PGI-S. The response quartiles were used to establish the recommended cut-off points for the outcome measure corresponding to “none” or “mild” vs “moderate” or “severe” categories on the PGI-S. When there was overlap or a gap in quartiles between adjacent PGI-S groups, half their differences were used.
4. Those with a PGI-S baseline score of mild, moderate, or severe
5. Those with a baseline score on the outcome measure corresponding to mild, moderate, or severe using ROC-derived cut-offs where the following approach was taken:

- A series of logistic regressions using the outcome measure score as the independent variable and a binary variable (“none” vs “mild” or “moderate” or “severe”) as the dependent variable. The optimal cut-off values for different levels of symptom severity were determined by maximizing Youden’s J statistic (the sum of sensitivity and specificity minus one) when the AUC exceeded 0.70.
6. Those with a baseline score on the outcome measure corresponding to mild, moderate, or severe using descriptive analysis where the following approach was taken:
- The distribution of the outcome measure score was categorized at baseline using the PGI-S. The response quartiles were used to establish the recommended cut-off points for the outcome measure corresponding to “none” vs “mild” or “moderate” or “severe” categories on the PGI-S. When there was overlap or a gap in quartiles between adjacent PGI-S groups, half their differences were used.

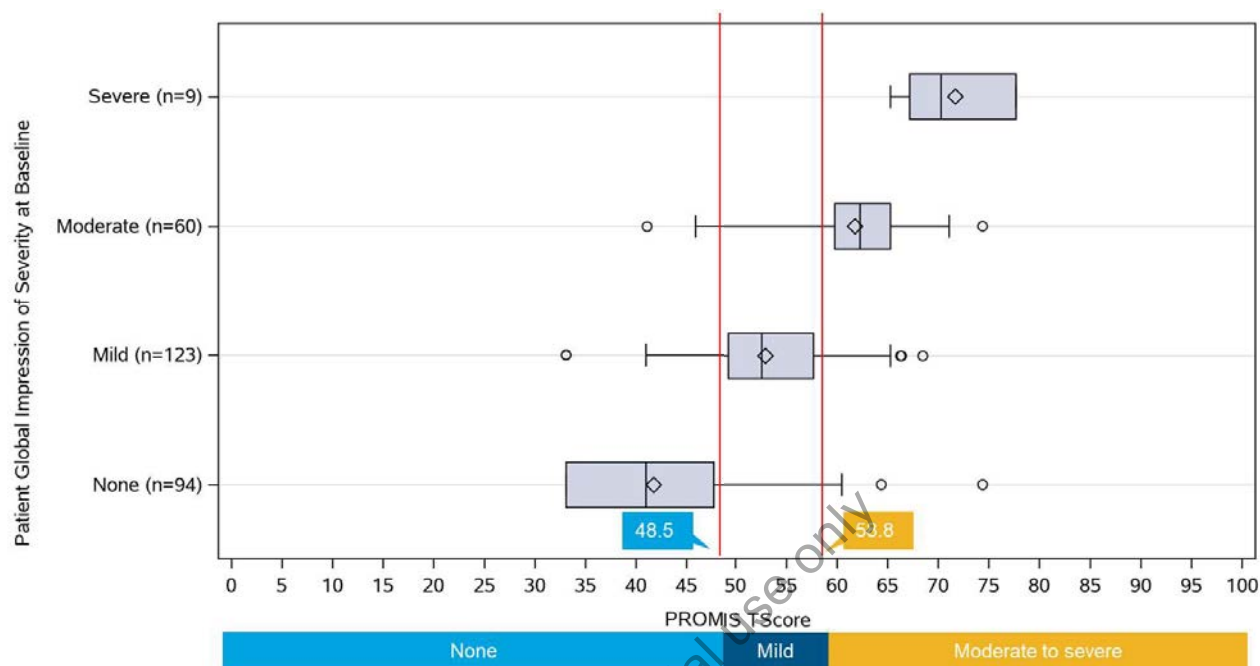
These separate approaches produced different baseline scores for the outcome measures and a variable number of patients in the “symptomatic” subgroups. Selection of the most appropriate approach to defining a subgroup that is “symptomatic at baseline” from the six approaches was based on the following criteria for the PROMIS Fatigue Short Form 8a total T-score, MFSAF TSS4, and the MPN-SAF problems with concentration symptom score:

- The subgroup should offer potential to detect improvement in the outcome measure (based on mean baseline score)
- The subgroup should be large enough to detect a treatment effect between rufertide and placebo in the outcome measure
- The subgroup should be defined based on approaches consistent with regulatory agency guidance

Defining a Symptomatic Subgroup for the PROMIS Fatigue Short Form 8a Total T-score

Protagonist proposes to use the descriptive analyses (i.e., approach #3) to define a symptomatic subgroup for the PROMIS Fatigue Short Form 8a total T-score endpoint (i.e., based on patients reporting moderate or severe symptoms on the PROMIS Fatigue Short Form 8a total T-score at baseline). The cut-off for the moderate-severe subgroup is 58.8 (see [Figure 2](#)). As such, a PV-focused PRO-related secondary endpoint will be “Mean change from baseline at end of Part 1a in the PROMIS Fatigue Short Form 8a total T-score among patients who have moderate or severe symptoms at baseline”, where a moderate or severe PROMIS Fatigue Short Form 8a total T-score at baseline is defined as a T-score ≥ 58.8 .

Figure 2: Severity Cut-off for the PROMIS Fatigue Short Form 8a Total T-score in the “Moderate to Severe” Subgroup of Patients Using the Descriptive Analysis of the Baseline PROMIS Fatigue Short Form 8a Total T-score

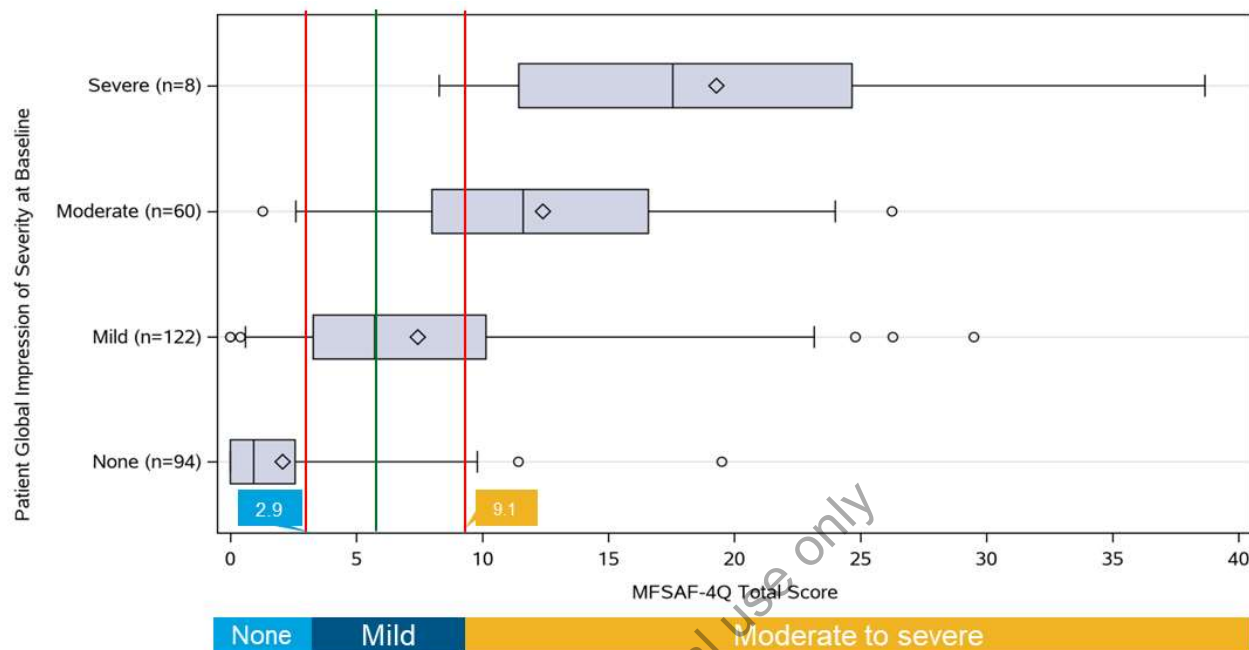


PROMIS Fatigue Short Form 8a=Patient-Reported Outcome Measurement Information System Fatigue Short Form 8a

Defining a Symptomatic Subgroup for the MFSAF TSS4

Protagonist proposes to use the descriptive analyses (i.e., approach #3) to define a symptomatic subgroup for the MFSAF TSS4 endpoint (i.e., based on patients reporting moderate or severe symptoms on the MFSAF TSS4 at baseline). The cut-off for the moderate-severe subgroup is 9.1 (see [Figure 3](#)). As such, a second PV-focused PRO-related secondary endpoint will be “Mean change from baseline at end of Part 1a in the MFSAF TSS4 among patients who have moderate or severe symptoms at baseline”, where a moderate or severe MFSAF TSS4 at baseline is defined as ≥ 9.1 .

Figure 3: Severity Cut-off for the MFSAF TSS4 in the “Moderate to Severe” Subgroup of Patients Using the Descriptive Analysis of the MFSAF TSS4 Baseline Score

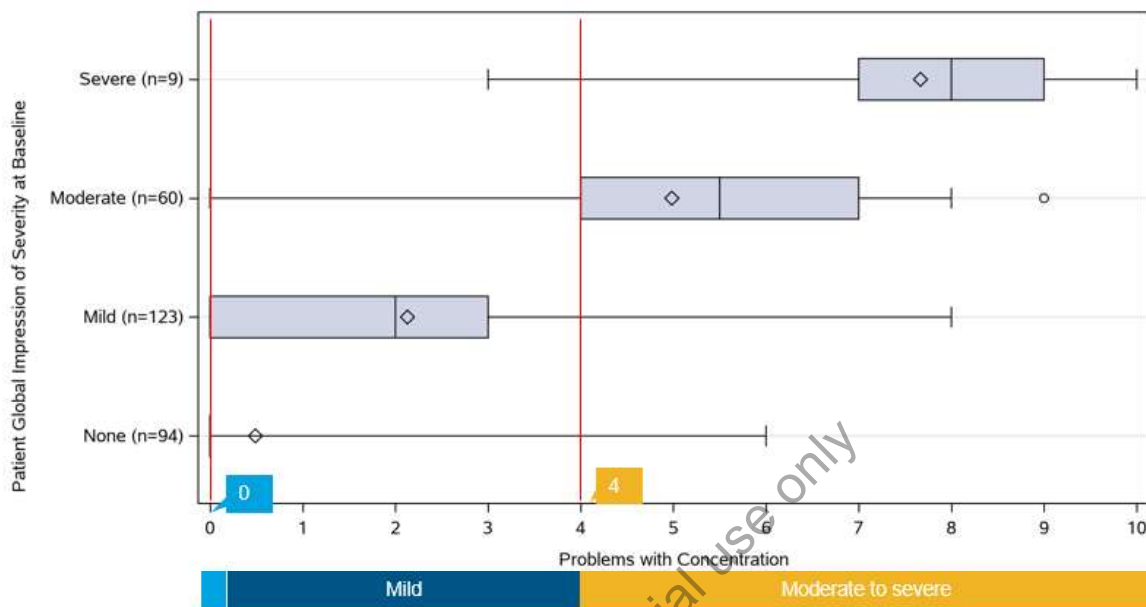


MFSAF TSS4= Myelofibrosis Symptom Assessment Form Total Symptom Score (4-items)

Defining a Symptomatic Subgroup for the MPN-SAF Problems with Concentration Score

Protagonist proposes to use the descriptive analyses (i.e., approach #3) to define a symptomatic subgroup for the MPN-SAF problems with concentration endpoint (i.e., based on patients reporting moderate or severe symptoms on the MPN-SAF problems with concentration score at baseline). The cut-off for the moderate-severe subgroup is 4 (see Figure 4). As such, a third PV-focused PRO-related secondary endpoint will be “Mean change from baseline at end of Part 1a in the MPN-SAF problems with concentration score among patients who have moderate or severe symptoms at baseline”, where a moderate or severe MPN-SAF problems with concentration at baseline is defined as a score ≥ 4 .

Figure 4: Severity Cut-off for the MPN-SAF Problems with Concentration Symptom Score in the “Moderate to Severe” Subgroup of Patients Using the Descriptive Analysis of the Baseline MPN-SAF Problems with Concentration Symptom Score



MPN-SAF=Myeloproliferative Neoplasm Symptom Assessment Form