Statistical Analysis Plan for Study M16-109

A Phase 2 Open-Label Study Evaluating Tolerability and Efficacy of Navitoclax Alone or in Combination with Ruxolitinib in Subjects with Myelofibrosis

Date: 31 January 2023

Version 4.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This statistical analysis plan (SAP) describes the statistical analyses to be conducted by the Data and Statistical Science Department for Navitoclax (ABT-263) Study M16-109. The study examines the efficacy and safety of Navitoclax alone or in combination with Ruxolitinib in subjects with primary Myelofibrosis (PMF) or secondary MF (Post Polycythemia Vera [PPV-MF] or Post Essential Thrombocythemia [PET-MF]).

This SAP provides details to further elaborate statistical methods as outlined in the study protocol (amendment 11). It will not be updated in case of administrative changes or amendments to the protocol unless if the changes impact the analysis. In addition to this SAP, Statistical Programming Plan (SPP) will document details to guide the statistical programming work.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The primary objective of the study is to:

- Evaluate the effect of navitoclax alone or in combination with ruxolitinib on spleen volume.

The secondary objectives of the study are:

- To assess the effect of navitoclax alone or in combination with ruxolitinib on total symptom score (TSS) as assessed by the Myelofibrosis Symptom Assessment Form (MFSAF) version 4.0 diary.
- To evaluate the effect of navitoclax alone or in combination with ruxolitinib on bone marrow fibrosis.

- To determine the rate of anemia response associated with navitoclax alone or in combination with ruxolitinib.
- To describe the safety profile and PK profile observed with navitoclax alone or in combination with ruxolitinib.

The exploratory objectives of the study are:

- To evaluate the effect of navitoclax alone or in combination with ruxolitinib on the onset, magnitude, and duration of disease response, including effects on spleen and anemia.
- To evaluate the effect of navitoclax alone or in combination with ruxolitinib on measures of health-related quality of life including total symptom score, fatigue, and physical functioning.
- To evaluate the effect of navitoclax alone or in combination with ruxolitinib on overall survival and progression free survival.
- Exploration of biomarkers predictive of navitoclax activity and response may be performed.

2.2 Study Design Overview

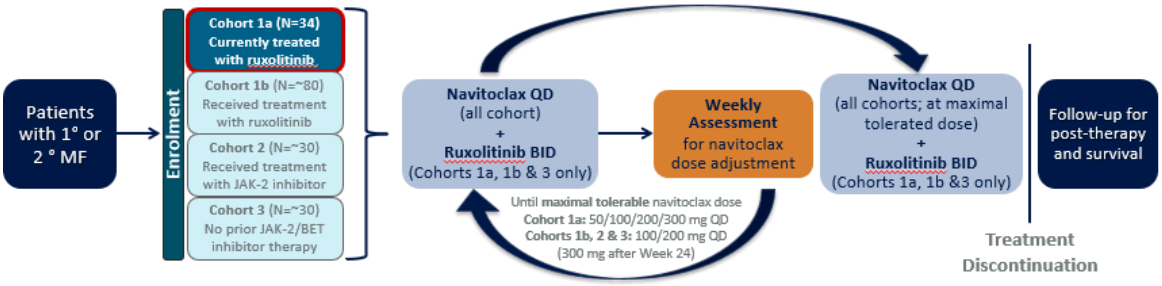
This is a Phase 2, multicenter, open-label study designed to evaluate the tolerability and efficacy of navitoclax alone or in combination with ruxolitinib in subjects with primary or secondary myelofibrosis (PPV-MF, PET-MF). Subjects enrolled in Cohort 1a must have received ruxolitinib therapy for at least 12 weeks and currently be on a stable dose of ≥ 10 mg BID of ruxolitinib. For Cohort 1b subjects must have received treatment with ruxolitinib for at least 24 weeks and currently on stable dose of ≥ 10 mg BID of ruxolitinib or had < 24 weeks of ruxolitinib with evidence of disease progression. For Cohort 2, subjects must have received prior treatment with a JAK-2 inhibitor. For Cohort 3, subjects must not have received prior treatment with a JAK-2 inhibitor, BH3 mimetic or BET inhibitor.

Approximately 174 subjects at approximately 135 sites globally will be enrolled in the following cohorts:

- Cohort 1a: 34 subjects
- Cohort 1b: approximately 80 subjects will be enrolled to ensure approximately 70 subjects with prior Ruxolitinib exposure.
- Cohort 2: approximately 30 subjects
- Cohort 3: approximately 30 subjects

Figure 1 is the study schema per protocol.

Figure 1. Study Schematic



Subjects will continue their treatment until end of clinical benefit, unacceptable toxicity or subject meets other protocol criteria for discontinuation (whichever occurs first). All subjects will have a treatment completion visit (TCV) performed when treatment is discontinued.

Response and progression will be assessed by the Investigator based on the International Working Group-Myeloproliferative Neoplasms Research and Treatment and European Leukemia Net (IWG-MRT/ELN).[1] Spleen volume will be assessed independently by a central imaging vendor and interpretations will be transmitted to investigators, as available, and will be utilized in determining response and progression.

2.3 Treatment Assignment and Blinding

This is an open-label and non-randomized study.

2.4 Sample Size Determination

The primary endpoint for each Cohort of the study is SVR_{35W24} (at least 35% reduction in spleen volume by MRI/CT at Week 24 from baseline). Approximately 174 subjects in total will be enrolled in this study across all Cohorts. Point estimate and exact 95% confidence interval of SVR_{35W24} rate corresponding to observed number of subjects with SVR_{35W24} are presented in Table 1 for each Cohort.

Table 1. Point Estimates and 95% CI of Rate of SVR_{35W24} Corresponding to Observed Number of Subjects with SVR_{35W24}

				Exact 95% CI		
Cohort	Sample size	Number of Subjects with SVR_{35W24}	Point Estimate (SVR_{35W24} Rate) (%)	Lower Limit (%)	Upper Limit (%)	Half Width of CI
1a	34	16	47.06	29.78	64.87	17.55
1a	34	18	52.94	35.13	70.22	17.55
1a	34	20	58.82	40.70	75.35	17.33
1a	34	22	64.71	46.49	80.25	16.88
1b	70	32	45.71	33.74	58.06	12.16
1b	70	34	48.57	36.44	60.83	12.19
1b	70	36	51.43	39.17	63.56	12.19
1b	70	38	54.29	41.94	66.26	12.16
2 or 3	30	12	40.00	22.66	59.40	18.37
2 or 3	30	14	46.67	28.34	65.67	18.67
2 or 3	30	16	53.33	34.33	71.66	18.67
2 or 3	30	18	60.00	40.60	77.34	18.37

The study with 34 subjects in Cohort 1a, 70 subjects in Cohort 1b, 30 subjects in Cohort 2 and 30 subjects in Cohort 3 showed in the above table, respectively, a reasonable precision for estimate of SVR_{35W24}. Also, if true probability of experiencing a serious

adverse event (SAE) due to the study treatment is 10%, then the probability of observing at least one SAE in 34 subjects is more than 97% in Cohort 1a; the probability of observing at least one SAE in 80 subjects is more than 99% in Cohort 1b; the probability of observing at least one SAE in 30 subjects is more than 95% in Cohort 2 and 3, respectively. Therefore, from safety assessment prospective the proposed sample size is adequate.

3.0 Endpoints

3.1 Primary Endpoint

SVR_{35W24}, at least 35% reduction in spleen volume at Week 24 from baseline as measured by MRI or CT according to the IWG criteria.[1]

3.2 Secondary Endpoint(s)

Secondary efficacy endpoints are as follows:

- TSS_{50W24}, at least 50% reduction in TSS at Week 24 from baseline as measured by MFSAF version 4.0[2]
- Anemia response
- Change in bone marrow fibrosis grade according to the European consensus grading system[3]

3.3 Additional Efficacy Endpoint(s)

Exploratory efficacy endpoints are as follows:

- SVR_{35}, at least 35% reduction in spleen volume from baseline as measured by MRI or CT according to the IWG criteria[1] at any time point during study
- Duration of SVR_{35}
- At least 50% reduction in palpable spleen length from baseline
- At least 50% reduction in TSS at any time point from baseline as measured by MFSAF v4.0

- Change in TSS and each MFSAF (v4.0) score from baseline
- Duration of anemia response
- Change in EORTC-QLQ-C30 score from baseline
- Change in PROMIS score from baseline
- Overall response per modified IWG criteria
- Progression Free Survival (PFS) per modified IWG criteria
- Overall survival (OS)
- 10%, 20%, and 30% Reduction in spleen volume from baseline as measured (SVR10, SVR20, and SVR30) by MRI/CT at any time point during study
- 10%, 20%, and 30% Reduction in spleen volume from baseline as measured (SVR10, SVR20, and SVR30) by MRI/CT at Week 24
- Post-baseline PRBC transfusion independence

3.4 Safety Endpoint(s)

Safety evaluations include the following:

- AE, Adverse event
- SAE, serious adverse event
- Deaths
- ECG, electrocardiogram variables
- Vital signs, clinical laboratory testing (hematology and clinical chemistry)
- Transfusion requirement for PRBC and platelets

3.5 Additional Endpoint(s)

- Biomarker endpoints as follows:
  - Change in Variant Allele Frequency
- Pharmacokinetic (PK) endpoints for navitoclax and ruxolitinib are as follows:
  1. C_{max}, maximum observed plasma concentration

2. T_{max}, the time to C_{max}
3. AUC_t, the area under the plasma concentration-time curve (AUC) from time 0 to the time of the last measurable concentration
4. β, the terminal phase elimination rate constant, only for navitoclax if the data warrants
5. AUC_∞, Area under the plasma concentration-time curve from time 0 to infinite time, only for navitoclax if the data warrants

Additional exploratory biomarker and PK endpoints may be evaluated (if required) and these will be analyzed without specifying in this study SAP.

4.0 Analysis Populations

In this study, study treatment is navitoclax + ruxolitinib (Cohort 1a, 1b and 3) or navitoclax alone (Cohort 2). Unless otherwise noted, study drug refers to navitoclax and ruxolitinib for combination therapy (Cohort 1a, 1b, and 3); and navitoclax for mono therapy (Cohort 2). Cohort 1 refers to Cohort 1a and Cohort 1b together.

The following sets will be used for analyses:

Full Analysis Set (FAS) includes all enrolled subjects who received at least one dose of navitoclax. The FAS will be used for all demographic/baseline analyses. Unless otherwise specified, the FAS is also the analysis population for the primary and the secondary endpoints.

The Safety Analysis Set includes all enrolled subjects who received at least one dose of navitoclax. The safety analysis set will be used for all safety analyses, unless otherwise specified.

The pharmacokinetics (PK) analysis set includes all the enrolled subjects who received at least one dose of navitoclax and have at least one post-dose PK assay results. The PK set will be used for PK analyses.

5.0 Subject Disposition

The FAS will be used to summarize subject disposition by cohort. The total number of subjects who were treated will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized by cohort.

- Subjects who were treated with at least one dose of navitoclax
- Subjects who discontinued study drug- navitoclax (all reasons and primary reason)
- Subjects who discontinued study drug- ruxolitinib (all reasons and primary reason)

6.0 Study Drug Duration and Compliance

6.1 General Considerations

Safety analysis set will be used for the analyses of study drug duration and compliance. Descriptive statistics (number of non-missing observations, mean, standard deviation, median, maximum, and minimum) will be provided for navitoclax and ruxolitinib, respectively, by Cohort, unless otherwise stated. Based on the dosing schedule, for Cohort 1 and 3, data from navitoclax and ruxolitinib will be separately summarized, and for Cohort 2 data from navitoclax will be summarized.

6.2 Duration of Study Drug

Duration of study drug is calculated for each subject by the following formula, for navitoclax and ruxolitinib, respectively:

Duration of study drug (days) = *Date of the last dose of study drug -Date of the first dose of study drug+1.*

The duration of study drug will be categorized as some interval of time (e.g., every 28-day). The number and percentage of subjects in each interval will be summarized.

6.3 Subject-Years of Exposure

Subject-years of study drug exposure is calculated as below for navitoclax and ruxolitinib, respectively:

$$\text{Subject-years of exposure (sub-yr)} = \frac{\text{Sum of duration of study drug for all subjects (days)}}{365.25}$$

6.4 Dose intensity (DI) and Relative Dose Intensity (RDI)

An observed dose intensity (ODI) will be derived for each subject who receives the study drug based on how it was actually administered. A planned dose intensity (PDI) will also be derived based on the protocol-specified schedule of the study drug.

For ODI it is calculated as:

$$\text{ODI} = \frac{\text{Actual total dose}}{\text{Actual total duration of treatment exposure (days)}}$$

For PDI it is calculated as:

$$\text{PDI} = \frac{\text{Planned total dose over protocol specified}}{\text{Actual total duration of treatment exposure (days)}}$$

RDI is defined as: RDI = (ODI/PDI)×100%.

The RDI will be summarized based on the study drug administration form in EDC. Ruxolitinib administered before the first dose of navitoclax is considered as prior study treatment therapy collected in the Study Cancer Prior Systemic Therapies (Myelofibrosis)

form in EDC for Cohort 1 (Cohorts 1a and 1b). Ruxolitinib administered after the last dose of navitoclax for Cohort 1 (Cohorts 1a and 1b) is considered as post-study treatment collected in Post Study Drug Systemic Therapies form in EDC.

The actual total duration of treatment exposure in the above equations is defined as the last dose date minus first dose date + 1 for navitoclax or ruxolitinib, respectively. Descriptive statistics will be presented for navitoclax (for all Cohorts) and ruxolitinib (for Cohort 3, Cohort 1a and Cohort 1b) respectively.

Number and percentage of subjects who have dose reductions or interruptions, will be summarized.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

7.1 General Considerations

Demographics, baseline/disease characteristics, medical history, and prior and concomitant medications will be performed on the FAS by Cohort.

Definition of First Date of and Last Date of Study Treatment

First date of study treatment is the date of the first dose of navitoclax. Last date of study treatment is the date of the last dose of navitoclax.

Definition of Baseline

In general, baseline value refers to the last non-missing measurement value collected up to the first date of study treatment, unless otherwise specified in Section 7.2.

7.2 Additional Baseline Definitions

7.2.1 Baseline TSS Value

The total symptom score (TSS) will be measured by Myelofibrosis Symptom Assessment Form (MFSAF) v4.0. The daily TSS scores will be calculated by summing up scores from 7 items of the MFSAF. All seven items must be completed for a daily TSS score to be computed. The worst (highest) daily TSS score will be considered as the daily TSS for a day when multiple responses on the same day were reported.

For subjects who enrolled under Study Protocol Amendment 7 or earlier version, the baseline TSS will be calculated by averaging all available daily TSS score collected up to the first date of study treatment (including the first date of study treatment).

For subjects who enrolled under Study Protocol Amendment 8 or later version, the baseline TSS score is defined as the average of the daily TSS score within 7 days before the first date of study treatment (not including the first date of study treatment). At least 4 days of scores are required to calculate the baseline TSS score. If a subject does not have at least 4 days of scores before the first date of the study treatment (not including the first date of study treatment), the subject will not have a baseline TSS score (see detail in Appendix C).

7.2.2 Baseline Hemoglobin Laboratory Value

For hemoglobin laboratory test the baseline value will be defined as follows:

For subjects who received PRBC transfusion (for any reason) within 4 Weeks on or before the first date of study treatment, the baseline value is the minimum value observed before the transfusion, and the last non-missing value on or prior to the first date of study treatment.

For subjects who do not receive PRBC transfusion within 4 weeks on or before the first date of study treatment, the baseline value is the last observed value on or before the first date of study treatment.

7.2.3 Baseline Renal Function

Baseline renal function will be classified based on the estimated creatinine clearance (CLcr) per Cockroft-Gault method,[4] The last non-missing serum creatine value collected on or before the first date of study treatment will be used as baseline (assuming Day 1 collection is prior to dose as timestamp was not collected). The following table illustrates the classification for renal function:

Table 2. Renal Function Group Based on CLcr

Renal Function Group	CLcr (mL/min)
Normal	CLcr ≥ 90
Mild Impairment	60 ≤ CLcr < 90
Moderate Impairment	30 ≤ CLcr < 60
Severe Impairment	15 ≤ CLcr < 30

7.2.4 Baseline Hepatic Function

Baseline hepatic function will be classified based on the National Cancer Institute Organ Dysfunction Working Group (NCI-ODWG) definition[5] using total bilirubin and aspartate transaminase (AST) values. The last observed bilirubin and AST values collected before the first date of study treatment (including Day 1 assuming lab parameter collected before study treatment on Day 1) will be used as baseline. The following table illustrates the classification of hepatic function:

Table 3. Hepatic Function Group Based on Total bilirubin and AST

Hepatic Function Group	Total Bilirubin (mg/dL)	AST (IU/L)
Normal	≤ ULN	≤ ULN
Mild Impairment	≤ ULN > ULN and ≤ 1.5 × ULN	> ULN any value
Moderate Impairment	> 1.5 × ULN and ≤ 3 × ULN	any value
Severe Impairment	> 3 × ULN	any value

Note: ULN are based on reported values from each lab.

7.2.5 DIPSS+ Score at Study Entry

For subjects with missing karyotype information at study entry, the DIPSS+ score at study entry will be calculated by utilizing the DIPSS risk score, platelet count, read blood cell transfusion need collected in EDC form 'DIPSS scoring – Study Entry' according to DIPSS+ scoring system in protocol Table 15 and by assuming these subjects do not have unfavorable karyotype.

7.3 Demographics and Baseline Characteristics

FAS will be used for this analysis. Categorical variables will be summarized with the number and percentage of subjects. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean, standard deviation, median, minimum, and maximum).

The following categorical demographic and baseline disease characteristics (but not limited to) will be summarized as shown in Table 4:

Table 4. Demographic and Baseline Characteristics (Categorical Variables)

Categorical variable (possible values)
Age (18 - < 65, 65 - < 75, ≥ 75)
Gender (Male, Female)
Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander)
Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
Region (US, Japan, Rest of World [ROW])
ECOG performance Status (0,1,2)
Baseline Renal Function (Normal, Mild, Moderate, Severe)
Baseline Hepatic Function (Normal, Mild, Moderate, Severe)
Type of MF (Primary MF, Secondary MF)
Prior lines of MF treatment (1,2, 3, > 3)
Response to prior ruxolitinib at screening (Refractory, Relapsed, Disease progression while on therapy, intolerance, stable disease, PV progressed to MF, ET prolonged to MF, not applicable, Other)

Table 4. Demographic and Baseline Characteristics (Categorical Variables) (Continued)

Categorical variable (possible values)
Reason for treatment change at study entry (Splenomegaly, Worsening symptoms, Increased need for transfusions, Anemia, Thrombocytopenia, Increasing WBC, Increasing LDH, Others)
Worsening symptoms as a reason for treatment change at study entry (Fatigue, Altered concentration, Early satiety, Inactivity, Night sweats, Itching, Bone pain, Weight loss, Fevers, Edema, Abdominal discomfort, Others)
Splenomegaly only as a reason for treatment change at study entry (Yes, No)
Worsening symptoms only as a reason for treatment change at study entry (Yes, No)
Worsening labs only as a reason for treatment change at study entry (Yes, No)
At least one worsening labs as a reason for treatment change at study entry (Yes, No)
Splenomegaly and worsening labs as a reason for treatment change at study entry (Yes, No)
Splenomegaly and worsening symptoms as a reason for treatment change at study entry (Yes, No)
Worsening symptoms and worsening labs as a reason for treatment change at study entry (Yes, No)
Splenomegaly, worsening symptoms, and worsening labs as a reason for treatment change at study entry (Yes, No)
Change from nadir in palpable spleen length with prior ruxolitinib/JAK2 inhibitor (< -50%, ≥ -50% to < -25%, ≥ -25% to < 0%, 0%, > 0% to < 25%, ≥ 25% to < 50%, ≥ 50%)
Transfusion dependence status based on Investigators (Dependent, Independent)
Mutation status at study entry from central review (JAK2, CALR, MPL, other mutation categories if applicable)
Platelet count (10^9/L) (< 75, ≥ 75 to < 100, ≥ 100)
Hemoglobin (g/dL) (< 8, ≥ 8 to < 10, ≥ 10)
Unfavorable karyotype at study entry (Yes, No, Not evaluable)
Dynamic international prognostic scoring system (DIPSS) risk at study entry (Low, Intermediate-1, Intermediate-2, High)
DIPSS risk at time of diagnosis (Low, Intermediate-1, Intermediate-2, High, Not evaluable)
DIPSS risk at time of ruxolitinib/Jak-2 inhibitor (Low, Intermediate-1, Intermediate-2, High, Not evaluable)
DIPSS plus risk at study entry (Low, Intermediate-1, Intermediate-2, High, Not evaluable)
Bone Marrow Fibrosis Grade at baseline (MF-0, MF-1, MF-2, MF-3)

Note: Worsening labs for prior treatment change includes increased need for transfusions, anemia, thrombocytopenia, increasing WBC or increasing LDH.

The following continuous baseline disease characteristics will be summarized as shown in Table 5.

Table 5. Demographic and Baseline Characteristics (Continuous Variables)

Continuous Variable
Age (year)
Body weight (kg)
Body height (cm)
BMI (kg/m^2)
Number of prior lines of MF treatments
Number of days from last prior ruxolitinib to study drug start
Time (months) since MF diagnosis
Duration (weeks) of prior ruxolitinib exposure
Spleen length by palpation (cm)
Spleen volume (cm^3)
Hemoglobin (g/dL)
Platelet count (10^9/L)
White blood cell (10^9/L)
Neutrophil count (10^9/L)
LDH (U/L)
Baseline TSS

7.4 Medical History

Medical history and ongoing conditions, including MF related conditions and symptoms will be summarized and listed. Separate summaries will be presented for ongoing and historical medical conditions. The summaries will be presented by primary system organ class (SOC) and preferred term (PT). Data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized. The SOC will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects

reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

7.5 Prior and Concomitant Medications

Prior and concomitant medications including prior treatments for MF and preceding treatment for polycythemia vera or essential thrombocythemia prior to developing secondary MF will be summarized by generic name. A prior medication is defined as any medication taken before the first date of study treatment.

A concomitant medication is defined as any medication that started before or after the first date of study treatment and continued to be taken after the first date of study treatment, but not after the last date of study treatment. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

Subjects who report the use of medications more than one time will be counted only once in the summary of any prior or concomitant medication. Subjects reporting two or more uses of the same medication will be counted only once in the total for the associated generic drug name.

8.0 Efficacy Analyses

8.1 General Considerations

The primary analysis population for primary endpoint is the Full Analysis Set. Unless otherwise specified, the secondary efficacy endpoints and exploratory efficacy endpoints will be analyzed in Full Analysis Set.

All efficacy results will be presented by cohort unless stated otherwise. For all analyses, unless otherwise specified, data collected after the cutoff date will be excluded from statistical analysis. All spleen volume related endpoints described in Section 3.0 will be based on the assessments from the central review of the radiographic images by an Independent Review Committee (IRC) as described in the protocol.

All response rates and the corresponding 95% confidence interval (CI) using the Clopper-Pearson method will be provided.

For the analyses of all time-to-event endpoints, survivorship functions will be estimated using Kaplan-Meier product-limit methodology. If reached, median time-to-event and its 95% CI will be presented.

For the analyses of all continuous endpoints, such as change from baseline in score, the scores will be summarized with baseline mean, visit mean, mean change from baseline, mean percentage change from baseline, standard deviation, and the corresponding 95% CIs.

Baseline measurement related to each efficacy endpoint for analysis has been described in Section 7.0.

The post-study treatment refers to Post Study Drug Systemic Therapies for Myelofibrosis, Post Study Drug Radiation Therapy and Post Study Drug Transplant. The disease progression (DP) refers to relapse or progressive disease in the Disease Response Assessment Form.

8.2 Handling of Missing Data

Missing data for all time-to-event endpoints will be handled by appropriate censoring in analysis and the censoring rule will be described in the following sub-sections.

Unless otherwise specified, for the analyses of all response rates, subjects without response will be considered as non-responders.

No missing data imputation will be applied for change from baseline in PRO domain score.

8.3 Primary Efficacy Endpoint(s) and Analyses

8.3.1 Primary Efficacy Endpoint

Spleen Volume Reduction ≥ 35% at Week 24 (SVR_{35W24}) per IWG Criteria

SVR_{35W24} rate is defined as the proportion of subjects who achieved at least 35% reduction from baseline spleen volume at Week 24. The analysis window for Week 24 is defined as Day 127 to Day 211 per study activity schedule. Data after the start of post-study treatment or disease progression (relapse and progressive disease) will be excluded from the analyses. Subjects who do not have percent change from baseline (missing baseline or missing post-baseline) of spleen volume in the analysis window will be considered as non-responders.

Primary analysis population for SVR_{35W24} is the Full Analysis Set (see Section 4.0 for the definition of Full Analysis Set).

8.3.2 Main Analysis of Primary Efficacy Endpoint

The SVR_{35W24} rate will be estimated, and the corresponding 95% confidence interval (Cl) will be provided using the Clopper-Person method.

8.4 Secondary Efficacy Analyses

8.4.1 Secondary Efficacy Endpoints

8.4.1.1 TSS Reduction ≥ 50% at Week 24 (TSS_{50W24})

TSS_{50W24} rate is defined as the proportion of subjects who achieved at least 50% reduction from baseline in TSS at Week 24. TSS score at Week 24 is defined as the average of the daily scores from a 7-day period closest to Week 24 nominal period. The Week 24 nominal period is from Day 163 to Day 169 for subjects enrolled before protocol amendment 8 and from Day 162 to Day 168 for subjects enrolled after protocol amendment 7. The detail of finding the 7-day period closest to the Week 24 nominal period from MFSAF data is as follows. Also refer to Appendix C for more details.

1. Step 1: Search TSS 7-day period data in Week 24 nominal period, Day 162-168 (after protocol amendment 7) or Day 163-169 (prior to protocol amendment 8).
2. Step 2: If Step 1 data are not available, extend the Week 24 nominal period in increments of ± 1 day (e.g., for subjects enrolled after protocol Amendment 7: Day 163-169 (+1 day expansion), Day 161-167 (-1 day expansion))
3. Repeat step 2 in increments of ±1 day until ±7 days from Week 24 nominal period if data from previous step are not available
4. The final TSS at Week-24 is the average of 7-day period closest to the Week 24 nominal period. The following convention is applied to the situation when a pair of 7-day period data has the same distance to Week 24 nominal period: select the 7-day period data which is before Week 24 nominal period.

Daily TSS will be considered as missing if not all responses for 7 items are provided. The worst (highest) daily TSS score will be considered the daily TSS for a day when multiple responses were reported on the same day. At least 4 days of scores in the 7-day period of Week 24 Visit are required to calculate the TSS score for Week 24. If a subject does not have at least 4 days of scores in the 7-day period of Week 24 Visit, the subject will not have a TSS score at Week 24. TSS scores obtained after disease progression or start of post-study treatment will not be included in the analysis.

Primary analysis population for TSS_{50W24} is Full Analysis Set. Subjects with missing percent change in TSS at Week 24 (e.g., zero TSS at baseline, or missing either baseline or post-baseline) are considered as non-responders of TSS_{50W24}.

8.4.1.2 Anemia Response

Anemia response rate is defined as the proportion of subjects who achieved anemia response.

Baseline hemoglobin mentioned below is defined in Section 7.2.2. The transfusion dependency status (transfusion independent and transfusion dependent) mentioned below

are based on investigators reporting in the Study Cancer History Form according PRBC transfusion events (≥ 6 units within 12 weeks and ≥ 1 unit within 28 days) prior to the start of study treatment.

For a subject who is transfusion independent (TI) at baseline with hemoglobin value < 10 g/dL, the subject will be considered as achieving anemia response if the post baseline hemoglobin level increases by ≥ 2 g/dL from baseline without receiving PRBC transfusion (for any reason) within 2 weeks and without any erythropoietin, erythropoietin mimetics within the last 4 weeks prior to the increase in hemoglobin level by ≥ 2g/dL was observed. Hemoglobin values after 30 days post last date of study treatment, the start of post-study treatment or disease progression whichever is earlier will not be considered in the analysis of anemia response.

For a subject who is transfusion dependent (TD) at baseline, anemia response is defined as a period of at least 12 consecutive weeks without PRBC transfusion after the first dose of study drug and on or prior to 30 days post last dose of study drug, the start of post-study treatment, disease progression or death, whichever occurs earlier.

Subjects who are baseline transfusion independent with baseline hemoglobin of ≥ 10 g/dL will not be included in the analysis since these subjects are not evaluable for anemia response.

Anemia response rate will also be summarized for all subjects who are evaluable for anemia response in the Full Analysis Set and by transfusion dependency status at baseline.

Missing post-baseline anemia response in Full Analysis Set applicable to evaluation of anemia response are considered as non-responders of anemia response.

8.4.1.3 Change in Bone Marrow Fibrosis Grade

Grade of bone marrow fibrosis was assessed by investigators according to the European consensus grading system. The proportion of subjects who achieve a reduction of at least 1 grade in bone marrow fibrosis from baseline at any time will be summarized. The bone

marrow fibrosis grade reduction of at least 1 grade at Week 24 will also be summarized. Data after post-study treatment or disease progression will be excluded from the analysis.

Subjects with missing bone marrow fibrosis grade (either baseline or post-baseline) or those with bone marrow fibrosis Grade 0 at baseline are excluded in the analysis for the change in bone marrow fibrosis grade.

Time to 1st achieving a reduction of at least 1 grade in bone marrow fibrosis from baseline will also be summarized. Mean and median time (week) of time to 1st achieving a reduction of at least 1 grade in bone marrow fibrosis from baseline will be provided.

8.4.2 Main Analyses of Secondary Efficacy Endpoints

TSS_{50W24} rate, anemia response rate and the bone marrow fibrosis grade reduction rate will be estimated. The corresponding 95% confidence interval (Cl) will be provided using Clopper-Person method.

Bone marrow fibrosis grades will be tabulated using a shift table with baseline values of Grade 0, Grade 1, Grade 2, Grade 3, or missing grade, versus minimum post-baseline of Grade 0, Grade 1, Grade 2, Grade 3, or missing grade.

8.5 Exploratory Efficacy Analyses

8.5.1 Exploratory Efficacy Endpoints

8.5.1.1 Spleen Volume Reduction ≥ 10%, ≥ 20%, 30%, and ≥ 35% at Any Time (SVR_{10}, SVR_{20}, SVR_{30} and SVR_{35}), Spleen Volume Reduction ≥ 10%, ≥ 20%, and ≥ 30% at Week 24

SVR_{10}, SVR_{20}, SVR_{30}, and SVR_{35} rates at any time are defined as the proportion of subjects who achieved at least 10%, 20%, 30%, and 35% reduction from baseline spleen volume any time, respectively. SVR_{10}, SVR_{20}, and SVR_{30} at Week 24 are defined as the proportion of subjects who achieved at least 10%, 20%, and 30% reduction from baseline spleen volume at Week 24, respectively. Data after the start of post-study treatment or disease progression will be excluded from these analyses. Subjects who do not have

percent change from baseline (missing baseline or missing post-baseline) in spleen volume will be treated as non-responders. In addition, time to first SVR ≥ 10%, first SVR ≥ 20%, first SVR ≥ 30%, and first SVR ≥ 35% will also be summarized.

Above exploratory endpoints will be analyzed in the Full Analysis Set.

8.5.1.2 Duration of SVR_{35}

Only subjects who achieved SVR_{35} will be included in this analysis for duration of SVR35. The Duration of SVR35 is defined as the number of days from the first date of SVR35 achievement to the first date of 1) SVR< 35% AND observing ≥ 25% spleen volume increase from nadir (the lowest spleen volume in the previous assessments) or 2) disease progression (relapse or progressive disease) per investigator by IWG/MRT criteria, whichever is earlier. Subjects without the defined event mentioned above will be censored at the last date of the following assessments on or before post-study treatment: MRI/CT scan for spleen volume assessment, bone marrow assessment, lab assessment, disease response assessment.

8.5.1.3 Reduction in Palpable Spleen Length

Rate of reduction in palpable spleen length is defined as the proportion of subjects who achieved at least 50% reduction from baseline in palpable spleen length any time during the study.

Subjects who had baseline spleen length of greater than 0 (> 0) will be included in this endpoint. Subjects who don't have post baseline spleen length will be excluded from this endpoint. Data after start of post-study treatment or disease progression will not be included in the analyses.

8.5.1.4 TSS Reduction ≥ 50% (TSS_{50})

TSS_{50} rate is defined as the proportion of subjects who achieved at least 50% reduction from baseline in TSS at any week during the study.

Subjects who do not have any post baseline TSS score will be treated as non-responders. Subjects with baseline TSS of 0 or without baseline TSS score will also considered as non-responders. Data after the start of post-study treatment or disease progression will not be included in the analyses.

TSS_{50} rate will be analyzed in the Full Analysis Set.

8.5.1.5 Change in MFSAF Score

Change in average daily MFSAF score of weekly interval (7-day) for each item and TSS (except for Week 23 and Week 25) from baseline will be summarized according to the scoring approach provided in Appendix C. Data after the start of post-study treatment, disease progression will not be included in the analyses. Change in MFSAF score will be analyzed among subjects who had baseline and post-baseline MFSAF score.

8.5.1.6 Post Baseline Red Blood Cell Transfusion

Post Baseline Red Blood Cell transfusion (for any reason) evaluation period is from after the first dose of study drug to earliest day of on or before the last dose of study drug + 30 days, or the start of post-study treatment.

Post baseline transfusion independent rate will be calculated as the proportion of subjects who achieved PRBC transfusion independent during the treatment period. PRBC Transfusion independent is defined as a period of at least 12 weeks (≥ 84 days) with no PRBC transfusions during the evaluation period. In addition, the rate of conversion will be calculated as proportion of subjects being post-baseline PRBC transfusion independent from baseline PRBC transfusion dependent.

The duration of PRBC transfusion independent will also be summarized for overall and by transfusion status at baseline (transfusion dependent and transfusion independent). The duration of transfusion independence is defined as the first time period that a subject received no PRBC transfusions for at least 12 weeks (≥ 84 days). The descriptive

statistics (median and range) will be provided for the duration of transfusion independence.

Incidence rate of PRBC transfusion/subject/month, defined as total number of PRBC transfusion from all subjects divided by total duration of treatment from all subjects (months).

Time from first date of study treatment to the first time receiving 2 or more units of PRBC transfusion will be summarized.

Above endpoints will be analyzed in the Full Analysis Set.

8.5.1.7 Duration of Anemia Response

Duration of anemia response will be summarized by baseline transfusion status.

For subjects who were transfusion independent at baseline and achieved anemia response, the evaluation period for duration of anemia response is from achieving anemia response to 30 days after last dose of study treatment. Duration of anemia response is defined as number of days from the first date of achieving anemia response described in Table 6 to the date of loss anemia response as described in Table 7.

Table 6. Date of First Anemia Response for Duration of Anemia Response

Transfusion Status at baseline	Date of First Response
TI	On the first date observing hemoglobin level increase by ≥ 2 g/dL from baseline without receiving PRBC transfusion within the last 2 weeks or any erythropoietin, erythropoietin mimetics within 4 weeks prior to the increased hemoglobin level
TD	On the first date observing no PRBC transfusion for at least 12 weeks

Table 7. Event and Censoring Condition for Duration of Anemia Response

Transfusion Status at baseline	Event/Censor	Date of Event/Censor
TI	Event: 1) decrease in hemoglobin level by 2 g/dL or more from the hemoglobin level at the first occurrence of achieving anemia response, or 2) need for PRBC transfusion before disease progression for the bleeding event* which was unrelated to navitoclax/ruxolitinib or 3) use of any erythropoietin, erythropoietin mimetics within 30 days of last dose of study drug	Earliest date of events
	Censor: if no above-mentioned events observed within 30 days of last dose of study drug	The last date of hemoglobin assessment after achieving anemia response, 30 days post last dose of study treatment, start of post-study treatment, disease progression, or death, whichever is earlier. Date of achieving anemia response will be the censor date if no hemoglobin assessment is obtained after achieving anemia response.

* Bleeding event is defined as a treatment emergent hemorrhage event from narrow search of Standardized MedDRA Queries (SMQ). PRBC occurring between the start of bleeding event and end/ongoing will be considered as need for PRBC transfusion for bleeding.

For subjects who were transfusion dependent at baseline, the duration of anemia response is defined as the 1st time period that a subject received no PRBC transfusions for at least 12 weeks (≥ 84 days). The descriptive statistics (median and range) will be provided for the duration of anemia response.

8.5.1.8 Overall Response Rate (ORR)

ORR is defined as the proportion of subjects who achieve best response of complete remission (CR) or partial remission (PR) per IWG criteria.[1] Subjects who do not have any disease assessment or who have started - post-study treatment prior to achieving best response of CR or PR will be considered as non-responders. Responses achieved after

relapse or progressive disease or starting post-study treatment will not be included. This endpoint will be analyzed in the Full Analysis Set.

8.5.1.9 Progression-Free Survival (PFS)

PFS is defined as time from the first date of study treatment to the date of first documented disease progression/relapse per IWG criteria as determined by the investigator, or death due to any cause, whichever is earlier. Disease progression and death after start of post-study treatment are not considered as an PFS event.

For subjects who do not experience a PFS event (disease progression, relapse, or death due to any cause) on or prior to the start of post-study treatment, PFS will be censored at the last adequate disease assessment (where IWG response evaluation is not missing) on or prior to the start of post-study treatment.

For subjects without any post baseline disease assessments who did not die, PFS will be censored at Day 1.

PFS will be analyzed in the Full Analysis Set.

8.5.1.10 Overall Survival (OS)

OS is defined as time from the first date of study treatment to the date of death due to any cause. If a subject has not died, then the data will be censored on the date when the subject was last known alive.

The last known alive date will be determined by selecting the last available date of the following study procedures: last dose date of study treatment, start date of adverse event, bone marrow collection, disease assessment, spleen assessment, vital signs assessment, electrocardiogram assessment, clinical laboratory collection (including COVID-19 test), study drug administration, start date of concomitant or post-treatment medicine or procedure, last known alive date in survival follow-up, biospecimen sample collection, quality of life assessments, and performance status.

OS will be analyzed in the Full Analysis Set.

8.5.1.11 Change in PROMIS Fatigue Score

Change in PROMIS fatigue score is defined as the change from baseline score by Visit. The scores are collected using PROMIS Fatigue SF7a. The summary score, symptom and impact scores will be calculated according to the scoring algorithm specified in Appendix C. Data after the start of post-study treatment, disease progression will not be included in the analyses. Change in PROMIS Fatigue Score will be analyzed among subjects who had baseline and post-baseline PROMIS Fatigue Score.

8.5.1.12 Change in EORTC QLQ-C30 Score

Change in EORTC QLQ-C30 score is defined as the change from baseline score by Visit. The scores are collected using EORTC QLQ-C30. The summary score of EORTC QLQ-C30 will be calculated according to the scoring algorithm specified in Appendix C. Data after the start of post-study treatment, disease progression will not be included in the analyses. Change in EORTC QLQ-C30 Score will be analyzed among subjects who had baseline and post-baseline EORTC QLQ-C30 score.

8.5.2 Main Analyses of Exploratory Efficacy Endpoints

The data will be analyzed per Section 8.1.

8.6 Additional Efficacy Analyses

8.6.1 Change in Variant Allele Frequency

Change from baseline in variant allele frequency any time during the post baseline period will be derived by specific driver gene, which is mutated-JAK2, -CALR, and -MPL, if baseline allele frequency and post-baseline allele frequency are available based on the Next Generation Sequencing (NGS) method in the peripheral blood. Data after the start of post-study treatment, disease progression WILL BE included in the analyses. Change in variant allele frequency will be analyzed among subjects who had baseline and post-baseline variant allele frequency.

Proportion of subjects achieving allele frequency reduction from baseline of ≥ 20% at Week 24 or any time during the study treatment will be summarized among subjects with baseline and post-baseline of variant allele frequency for Cohort 1a, Cohort 1b and Cohort 3. Maximum percentage reduction from baseline in allele frequency will be presented by the waterfall plot for all cohorts.

8.6.2 Main Analyses of Additional Endpoints

Descriptive statistics with visit mean, mean change from baseline, mean percentage change from baseline, standard deviation and range will be provided.

8.7 Efficacy Subgroup Analyses

Subgroup analyses will be performed for the primary SVR_{35W24} endpoint.

The following subgroups will be used (but not limited to):

- type of MF (Primary MF and Secondary MF)
- DIPSS risk at study entry (Inetermediate-1, Intermediate-2 and High risk).
- Age (18 - < 65 years, ≥ 65 years)

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized using the Safety Analysis Set defined in Section 4.0. The safety summaries will be presented by Cohort. For all safety analyses, unless otherwise specified, data collected up to the analysis data cutoff date will be used.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical

study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs (TEAE) are defined as any AE with the onset that is on or after the first dose of navitoclax until 30 days after the last date of study drug. Events where the onset date is the same as the study drug start date are assumed to be treatment emergent. If an incomplete or missing onset date was collected for an AE, the AE will be assumed to be treatment-emergent unless there is evidence to the contrary (e.g., the AE end date was prior to the date of the first dose of study drug). All TEAEs will be summarized overall, as well as by primary MedDRA SOC and PT. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing TEAEs will be summarized.

9.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any TEAE
- Any grade 3 TEAE
- Any grade 4 TEAE
- Any TEAE with grade 3, or 4
- Any TEAE with grade 3, 4 or 5
- Any serious TEAE
- Any TEAE with reasonable possibility related to navitoclax assessed by the investigator

- Any TEAE with reasonable possibility related to ruxolitinib assessed by the investigator
- Any TEAE leading to discontinuation of navitoclax
- Any TEAE leading to discontinuation of ruxolitinib
- Any TEAE leading to interruption of navitoclax
- Any TEAE leading to interruption of ruxolitinib
- Any TEAE leading to dose reduction of navitoclax
- Any TEAE leading to dose reduction of ruxolitinib
- Any TEAE leading to death
- All deaths
- Deaths occurring ≤ 30 days after last dose of navitoclax
- Deaths occurring > 30 days after last dose of navitoclax

9.2.3 Treatment-Emergent Adverse Events by MedDRA SOC and/or PT

TEAEs listed in Section 9.2.2 will be summarized by MedDRA SOC and PT; Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest grade and level of relationship to investigational product will be reported.

In addition, TEAEs will be summarized by PT and sorted by decreasing frequency for Cohort 1, and summarized by MedDRA SOC, PT and maximum NCI CTCAE grade.

9.2.4 SAEs (including deaths) and Adverse Events Leading to Death

SAEs (including deaths) and AEs leading to death will be summarized by MedDRA SOC and PT and in listing format.

9.2.5 Safety Topics of Interest

Safety topics of interest will be identified using the search criteria in Appendix E.

Selected AEs will be summarized by SOC and PT. The same overview summary as for TEAEs and a listing of subjects' data will be provided for each search.

In addition to the searches listed in Appendix E, the torsade de pointes/QT prolongation search will be provided. Following analysis may be provided using SMQ Torsade de Pointes/QT Prolongation (Broad) search criteria: The incidence of the events by preferred term, the same overview summary as for TEAEs and a listing of subjects' data.

9.2.5.1 Time to Onset

Time to onset of the safety topics of interest will be summarized using summary statistics (N, mean, standard deviation, median, minimum, and maximum). Time to onset is defined as the time from first dose of Navitoclax to the first occurrence of the event. Time to onset will be summarized for each defined event as well as first Grade ≥ 3 event. The distribution of time to onset will be summarized for the time period as described in Section 6.2. The time periods may be adjusted to account for sparse data in later time periods if necessary.

Additionally, summary statistics (number of subjects, mean, standard deviation, median, minimum, and maximum) will be generated for time (in days) to first TEAE leading to dose reduction and time (in days) to first TEAE leading to dose interruption.

9.2.5.2 Incidence and Prevalence by Time

Incidence and prevalence for each PT within the search and overall incidence may be provided for the time periods based on time from Day 1 of the study treatment to the occurrence of the event described in Section 6.2. The 30 days TEAE window will be applied and grouped into the last time period for each subject. The time periods may be adjusted to account for sparse data in later time periods if necessary.

For each time period, the incidence for any given PT will be calculated as follows:

- The numerator for each time period will be the number of subjects who had the first occurrence of an adverse event in that time period
- The denominator for each time period will be the number of subjects who took at least one dose of navitoclax in the time period and did not experience an event within any previous period
- Incidence for each time period = numerator/denominator.

For each time period, the prevalence for any given PT will be calculated as follows:

- The numerator for each time period will be the number of subjects who had an occurrence of an adverse event in that time period or in a previous time period and that was ongoing in the current time period
- The denominator for each time period will be the number of subjects who took at least one dose of navitoclax in the time period.
- Prevalence for each time period = numerator/denominator.

For each TEAE search, a plot capturing the incidence and prevalence (by time period) will be displayed.

9.2.5.3 Transfusion Requirements During Study Treatment

Number of subjects who received at least a total of two units (total number of transfusion units) of PRBC transfusion or any Platelet transfusions after first dose of Navitoclax to 30 days after last dose of Navitoclax will be summarized by Transfusion type (platelet, PRBC). The Total number of transfusions (Sum of number of transfusion) among all subjects will be summarized.

9.3 Analysis of Laboratory Data

9.3.1 Analysis of Shift from Baseline

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analysis of laboratory data. Data collected after 30 days of last dose of study treatment will be excluded in the analysis of laboratory data.

For shifts relative to National Cancer Institute Common Toxicity Criteria for Adverse Events, baseline and Maximum post-baseline laboratory observations will be categorized as grade 0, grade 1, grade 2, grade 3, or grade 4 according to NCI CTCAE grading. The baseline grade is defined as the grade of the last measurement collected on or prior to the first dose of study treatment (except baseline hemoglobin – See Section 7.2.2). In cases where multiple values are collected on the same day, the minimum grade and maximum grade value will be selected as the baseline grade and postbaseline grade, respectively, for that day. The maximum NCI toxicity grade value is the value with highest NCI toxicity grade collected after the first dose of study treatment.

For each graded laboratory test, shift tables will be generated that cross tabulate the number of subjects with baseline values of grade 0, grade 1, grade 2, grade 3, or grade 4 versus maximum or final values of grade 0, grade 1, grade 2, grade 3, or grade 4.

The separate laboratory shifts tables from baseline to the maximum value post baseline based on the two criteria below will be generated for each laboratory tests related to CTCAE:

- Shifts from Grade 0 (Normal) at baseline to Grade 1 - 4 Post-baseline (maximum), or worsening from an abnormal baseline value of at least one grade post- baseline (maximum).
- Shifts from Grade 0 - 2 at baseline to Grade 3 or 4 Post-baseline (maximum) and from grade 3 at baseline value to Grade 4 post-baseline (maximum).

For above shift tables, baseline grade of 0 (normal) will be imputed for all subjects with at least one post-baseline but missing a baseline value for each lab test. In above shift table subjects with a grade 4 baseline value will not be included.

A listing of all relevant laboratory parameters will be provided for each subject who has Grade 3 or higher. All measurements collected, regardless of the number of days after the last dose of study treatment, will be included in these listings.

9.3.2 Analysis of Key Laboratory Parameters

For key lab parameters platelet count, lymphocytes, neutrophils, hemoglobin, total bilirubin, ALT, AST, and alkaline phosphatase, box plots of laboratory values will be displayed at baseline, 1, 4, 8, 12, 16, 20, 24, 36, 48, 60, 72, 84, 96 weeks for each cohort.

9.4 Assessment of Potential Drug-induced Liver Injury

Elevations relative to the upper limit of normal (ULN) in alanine transaminase (ALT), AST, total bilirubin, and alkaline phosphatase as outlined in the FDA Guidance for Industry[6] pertaining to premarketing clinical evaluations for drug-induced liver injury (DILI) will be summarized using the maximum post baseline values. Laboratory values collected no more than 30 days following the last date of study treatment will be used in the analysis.

The number and percentage of subjects in each Cohort who have at least one observed post baseline value meeting the following criteria will be tabulated:

- ALT: > 3 ×, > 5 ×, > 10 ×, or > 20 × ULN
- AST: > 3 ×, > 5 ×, > 10 ×, or > 20 × ULN
- Total bilirubin > 1.5 ×, > 2 × ULN
- Alkaline phosphatase > 1.5 × ULN
- ALT and/or AST > 3 × ULN and total bilirubin > 1.5 × ULN at any day
- ALT and/or AST > 3 × ULN and total bilirubin > 2 × ULN at any day

- ALT and/or AST > 3 × ULN and total bilirubin > 2 × ULN within 72 hours (3 days)
- ALT and/or AST > 3 × ULN and total bilirubin > 2 × ULN at the same day

An evaluation of Drug Induced Serious Hepatotoxicity (eDISH) plot of the maximum post-baseline ALT value (as a multiple of ULN) vs. the maximum post baseline total bilirubin value (as a multiple of ULN), not necessarily concurrent, will also be utilized to assess for potential hepatotoxicity. Reference lines will be included at 3 × ULN for ALT and at 2 × ULN for total bilirubin. A similar eDISH plot will be presented for AST vs. total bilirubin.

A listing of liver function tests for subjects with observed ALT and/or AST (> 3 × ULN) accompanied by total bilirubin (> 2 × ULN) at any time on study will be provided.

9.5 Analysis of Vital Signs

Change from baseline in vital signs will be summarized. Vital signs values collected no more than 30 days following the last date of study treatment will be used in the analysis.

9.6 Analysis of ECG Parameters

A listing will be prepared that will include details about all abnormal ECG findings.

9.7 Safety Subgroup Analyses

The incidence of treatment-emergent AEs overview, by SOC and PT, and selected AEs will be assessed for the subgroups defined below:

- Type of MF (PMF, SMF)
- Age (18 - < 65 years, ≥ 65 years)

10.0 Other Analyses

10.1 PK Endpoints

PK samples will be collected following the schedule in protocol Table 7 and Table 8 and will be analyzed for navitoclax and ruxolitinib concentrations. The PK parameters of maximum observed plasma concentration (C_{max}), the time to C_{max} (peak time, T_{max}), and the area under the plasma concentration-time curve (AUC) from time 0 to the time of the last measurable concentration (AUC_t) will be determined using non-compartmental methods for intensive PK visits (Cohort 1a Week 0 Day 1 for navitoclax and ruxolitinib; drug-drug interaction sub cohort within Cohort 1b Week 0 Day 0 for ruxolitinib and Week 0 Day 7 for navitoclax and ruxolitinib; Cohort 2 Week 0 Day 1 for navitoclax). Other parameters for navitoclax, such as the terminal phase elimination rate constant (β) and the area under the plasma concentration-time curve from time 0 to infinite time (AUC_∞) will be determined if the data warrants.

10.2 Main Analyses of PK Endpoints

The analyses of PK endpoints are summarized in following:

- Tabulation and Summary Statistics

Plasma concentrations of navitoclax and ruxolitinib and their available pharmacokinetic parameter values will be tabulated by Cohort and visit. Summary statistics will be computed for each sampling time and each parameter. Significant pharmacokinetic sample time deviations will be identified and listed.

- Cohort 1b Sub-Study: Assess the Effect of Navitoclax on Ruxolitinib PK

A repeated measures model will be performed to assess the effect of navitoclax on ruxolitinib PK for the following ruxolitinib pharmacokinetic parameters: the natural logarithms of C_{max}, AUC_t and AUC_{inf} (if appropriate). The model will include regimen

(navitoclax in combination with ruxolitinib and ruxolitinib alone) as a fixed effect and each subject will be considered as a random effect.

The bioavailability of ruxolitinib in combination with navitoclax relative to ruxolitinib alone will be estimated. Point estimates and corresponding 90% CIs for the ratio of ruxolitinib in combination with navitoclax to ruxolitinib alone will be provided. Additional analyses may be performed if useful and appropriate.

11.0 Interim Analyses

There is no planned interim analysis. However, when all subjects from each Cohort complete Week 24 disease assessment, or discontinue from study, whichever occurs first, primary efficacy analysis will be performed for the corresponding Cohort along with summaries for efficacy and safety data from other cohorts.

11.1 Data Monitoring Committee

The study does not have an independent data monitoring committee (DMC), therefore safety or efficacy data will be reviewed by the Sponsor.

12.0 Overall Type-I Error Control

There is no formal multiplicity testing will be performed to control type-I error rate in this study.

13.0 Version History

Table 8. SAP Version History Summary

Version	Date	Summary
1.0	24 Nov 2020	Original version
1.1	16 Jun 2021	1. Clarify exclusion of data beyond new anti MF therapy, PD in primary/secondary efficacy analyses. 2. Clarify BL TSS definition. 3. Clarify BL Hemoglobin calculation. 4. Update the definition of anemia response and corresponding duration. 5. Clarify missing data handling in PRO data. 6. Add regression analyses of SVR_{35W24} vs baseline DIPSS prognostic variables. 7. Add subgroup analyses by age group. 8. Add subgroup analyses for DIPSS INT-1 pts in Cohort 1. 9. Clarify TEAE criteria for AEs on Day 1 and AEs with missing/incomplete dates. 10. Add AESI overview summary. 11. Add AE leading to death summary. 12. Add PK set. 13. Add clinically notable QT summary. 14. Add blood pressure shift summary. 15. Update safety topics of interest table. 16. Clarify sample size per protocol amendment. 17. Other editorial changes.
2.0	27 Sep 2021	1. Efficacy evaluable set added 2. SVR_{35W24} analysis window extended to accommodate schedule challenges due to COVID19. 3. Restrict data up to 30 days post last dose of study drug for anemia response and duration of anemia response. 4. Modification of duration of anemia response derivation. 5. Modify the missing data handling for spleen length to be consistent with SVR. 6. Other editorial changes.
2.1	18 Oct 2021	1. Modify the definition of duration of SVR35 to incorporate other objective measure which may indicates the end of response. 2. Update safety topics of interest table.

Version	Date	Summary
3.0 draft	08 May 2022	1. Modify anemia response, duration of anemia response 2. Modify PFS and OS 3. Remove analysis of vital signs 4. Modify PRBC transfusion analysis 5. Modify TSS Week 24 derivation
3.0	09 Jan 2023	1. Fix typo in version 3 draft 2. Efficacy evaluable set removed
4.0	31 Jan 2023	1. Add DIPSS+ score at study entry

14.0 References

1. Tefferi A, Cervantes F, Mesa R, et al. Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. Blood. 2013;122(8):1395-8.
2. Mesa RA, Schwager S, Radia D, et al. The Myelofibrosis Symptom Assessment Form (MFSAF): an evidence-based brief inventory to measure quality of life and symptomatic response to treatment in myelofibrosis. Leuk Res. 2009;33(9):1199-203.
3. Thiele J, Kvasnicka HM, Facchetti F, et al. European consensus on grading bone marrow fibrosis and assessment of cellularity. Haematologica. 2005;90(8):1128-32.
4. Guidance for Industry: Pharmacokinetics in Patients with Impaired Renal Function – Study Design, Data Analysis, and Impact on Dosing and Labeling. US Food and Drug Administration Center for Drug Evaluation and Research. March 2010.
5. Ramalingam SS, Kummar S, Saranopoulos J, et al. Phase I study of vorinostat in patients with advanced solid tumors and hepatic dysfunction: a National Cancer Institute Organ Dysfunction Working Group Study. J Clin Oncol. 2010;28(29):4507-12.

6. Guidance for Industry: Drug-Induced Liver Injury: Premarketing Clinical Evaluation. US Food and Drug Administration Center for Drug Evaluation and Research. July 2009.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study treatment.
- Subject took prohibited concomitant medication.

Appendix B. IWG/MRT/ELN Response Criteria for Myelofibrosis[1]

Table B-1. Summary of Response Criteria

Response Category	Response Requirements		
	Bone Marrow	**Peripheral Blood**	**Clinical**
Complete Remission (CR)	• Age-adjusted normocellularity • < 5% blasts • Grade ≤ 1 MF	• Hgb ≥ 10 g/dL • Neutrophils ≥ 1×10^9/L • Platelets ≥ 100×10^9/L • < 2% immature myeloid cells	• Resolution of disease symptoms • Spleen/Liver not palpable • No evidence of EMH
Partial Remission (PR)	Not Applicable	• Hgb ≥ 10 g/dL • Neutrophils ≥ 1×10^9/L • Platelets ≥ 100×10^9/L • < 2% immature myeloid cells	• Resolution of disease symptoms • Spleen/Liver not palpable • No evidence of EMH
	OR		
	• Age-adjusted normocellularity • < 5% blasts • Grade ≤ 1 MF	• Hgb ≥ 8.5 g/dL • Neutrophils ≥ 1×10^9/L • Platelets ≥ 50×10^9/L • < 2% immature myeloid cells	• Resolution of disease symptoms • Spleen/Liver not palpable • No evidence of EMH
Clinical Improvement	The achievement of anemia, spleen or symptom response without progressive disease or increase in severity of anemia, thrombocytopenia or neutropenia*		
Anemia Response	Transfusion-independent (baseline Hgb < 10 g/dL): Hgb increase ≥ 2 g/dL Transfusion dependent: becoming transfusion independent**		
Spleen Response	A baseline splenomegaly that is palpable at 5 – 10 cm below the LCM, becomes not palpable***A baseline splenomegaly that is palpable at > 10 cm below the LCM, decreases by ≥ 50%***MRI showing ≥ 35% SVR		
Symptoms Response	≥ 50% reduction in MF-SAF TSS		
Stable Disease	Belonging to none of the above response categories		

EMH = extramedullary hematopoiesis; LCM = left costal margin

* Increase in severity of anemia constitutes the occurrence of new transfusion dependency or a ≥ 2 g/dL decrease in Hgb from pretreatment baseline that lasts for ≥ 12 weeks. Increase in severity of thrombocytopenia or neutropenia is defined as a 2-grade decline, from pretreatment baseline, in platelet count or ANC. In addition, assignment to Clinical Improvement requires a minimum platelet count of ≥ 25×10^9/L and ANC ≥ 0.5×10^9/L.

** Transfusion dependency before study enrollment is defined as transfusions of at least 6 units of packed red blood cells (PRBC), in the 12 weeks prior to study enrollment. In addition, the most recent transfusion episode must have occurred in the 28 days prior to study enrollment. Response in transfusion-dependent patients requires absence of any PRBC transfusions during any consecutive "rolling" 12-week interval during the treatment phase.

*** Requires confirmation by MRI with a ≥ 35% SVR.

Table B-2. Summary of Cytogenetic Remission

	Cytogenetic Remission Requirements
Cytogenetic Remission Category	At least 10 metaphases must be analyzed Requires confirmation by repeat testing within 6 months
Complete Cytogenetic Remission	Eradication of a preexisting abnormality
Partial Cytogenetic Remission	≥ 50% reduction in abnormal metaphases (to qualify for PR: must have ≥ 10 abnormal metaphases at baseline)

Table B-3. Summary of Molecular Remission

	Molecular Remission Requirements
Molecular Remission Category	Must be analyzed in peripheral blood granulocytes Requires confirmation by repeat testing within 6 months
Complete Molecular Remission	Eradication of a pre-existing abnormality
Partial Molecular Remission	≥ 50% decrease in allele burden (to qualify for PR: must have ≥ 20% mutant allele burden at baseline)

Table B-4. Summary of Progression/Relapse Criteria

Progression/Relapse Category	Progression/Relapse Criteria
Progressive Disease	At least 1 of the following: • Appearance of a new splenomegaly that is palpable at least 5 cm below the left costal margin (LCM)* • A ≥ 100% increase in palpable distance below LCM, for baseline splenomegaly of 5 – 10 cm* • A 50% increase in palpable distance below LCM, for baseline splenomegaly of > 10 cm* • MRI showing a ≥ 25% increase in spleen volume from baseline • Leukemic transformation confirmed by a bone marrow blast count of ≥ 20% • A peripheral blood blast content of ≥ 20% associated with an absolute blast count of ≥ 1×10^9/L that lasts for at least 2 weeks
Relapse	At least 1 of the following: • No longer meeting criteria for at least Clinical Improvement after achieving Complete Response, Partial Response or Clinical Improvement • Loss of anemia response persisting for ≥ 1 month • Loss of spleen response persisting for ≥ 1 month
Cytogenetic/Molecular Relapse	Re-emergence of a pre-existing cytogenetic or molecular abnormality that is confirmed by repeat testing

LCM = left costal margin

* Requires confirmation by MRI with a ≥ 25% increase in spleen volume from baseline.

Appendix C. Patient-Reported Outcome (PRO) Assessments

MFSAF Version 4.0

MFSAF v4.0 is a symptom diary in which patients rate the following seven MF symptoms: fatigue, night sweats, abdominal discomfort, pruritus, pain under the ribs on the left side, early satiety, and bone pain on a daily basis using a scale from 0 (absent) to 10 (worst imaginable). The TSS ranges from 0 to 70 and is calculated by summing up the scores from the following 7 questions. The detailed analysis window will be provided in the SPP.

- How Severe Was Worst Bone Pain
- How Severe Was Worst Fatigue
- How Severe Was Worst Itching (Pruritus)
- How Severe Was Worst Night Sweats
- How Severe Was Worst Abdominal Discomfort
- Worst Feel Fullness After Beginning Eat (early satiety)
- Worst Pain Under Ribs Left Side

For each subject, the score will be calculated at each day of a week, and the weekly score is calculated by averaging the total scores across all available days in the week. It is important to note that, at least 4 days of scores are required to calculate the weekly score. However, for baseline score for subjects enrolled under Amendment 7 or before, at least one day of scores are required.

A total of 15 of 7-day period TSS (Table C-1) can be considered as Week 24 TSS. The final TSS at Week-24 is the average of 7-day period closest to the Week 24 nominal period.

Table C-1. Example of Week 24 Window for Subjects Enrolled after Protocol Amendment 7

RX Day	1	2	3	4	5	6	7	8 Week 24 Nominal Period	9	10	11	12	13	14	15
155	x														
156	x	x													
157	x	x	x												
158	x	x	x	x											
159	x	x	x	x	x										
160	x	x	x	x	x	x									
161	x	x	x	x	x	x	x								
162		x	x	x	x	x	x	x							
163			x	x	x	x	x	x	x						
164				x	x	x	x	x	x	x					
165					x	x	x	x	x	x	x				
166						x	x	x	x	x	x	x			
167							x	x	x	x	x	x	x		
168								x	x	x	x	x	x	x	
169									x	x	x	x	x	x	x
170										x	x	x	x	x	x
171											x	x	x	x	x
172												x	x	x	x
173													x	x	x
174														x	x
175															x

EORTC QLQ-C30

HRQoL and symptoms will be assessed with the EORTC-QLQ-C30 version 3 published in 2001. The QLQ-C30 is 30-item subject self-administered questionnaire composed of both multi-item and single-item measures on a four-point scale, with 1 as "not at all" and 4 as "very much." The scoring algorithms for scales will follow the scoring manual, except for the summary score defined below.

The EORTC QLQ-C30 Summary Score is calculated by averaging all calculated domain scores excluding scores from the domains of Global Health Status and Financial Difficulties. The summary score will be calculated if all of the required 13-domain scores are available. Symptom scales will be subtracted from 100 before averaging for summary score calculation.

Patient Global Impression of Change

The Patient Global Impression of Change (PGIC) scale will be utilized to assess patients' perceptions of change in their myelofibrosis symptoms over time. Patients during study visits will answer the following question: "Since the start of the treatment received in this study, your myelofibrosis symptoms are (1) Very much improved, (2) Much improved, (3) Minimally improved, (4) No change, (5) Minimally worse, (6) Much worse, (7) Very much worse."

PROMIS Fatigue Short Form (SF) 7a

PROMIS® is a system of highly reliable, precise measures of patient-reported health status for physical, mental, and social well-being. PROMIS instruments measure concepts such as pain, fatigue, physical function, depression, anxiety and social function. Fatigue will be assessed using the PROMIS Fatigue SF 7a that has been developed and validated for use in oncology populations. The PROMIS Fatigue SF 7a is a seven-item questionnaire that assesses the impact and experience of fatigue over the past 7 days. The recommended minimum important difference range is 3 – 5 points; the lower bound (3) is being used as the minimum important difference in this study. Five response options for

question 1-6 are: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Always, whereas question 7 has reversed response options.

Overall total score for PROMIS Fatigue is derived from raw scores according to the scoring manual of PROMIS Fatigue SF 7a. The symptom sub-domain of the PROMIS Fatigue SF 7a includes items 1-3 and the Impact domain includes items 4-7. The scoring for each sub-domain is calculated as the averaging the raw scores.

Appendix D. Bone Marrow: Grading Myelofibrosis

Table D-1. Consensus on Grading of MF

Grading	Description
MF-0	Scattered linear reticulin with no intersections (cross-overs) corresponding to normal bone marrow
MF-1	Loose network of reticulin with many intersections, especially in perivascular areas
MF-2	Diffuse and dense increase in reticulin with extensive intersections, occasionally with only focal bundles of collagen and/or focal osteosclerosis
MF-3	Diffuse and dense increase in reticulin with extensive intersections with coarse bundles of collagen, often associated with significant osteosclerosis.

Appendix E. Safety Topics of Interest

Safety Topics of Interest	Type of MedDRA Query	Broad or Narrow Search	Search Criteria
Thrombocytopenia	PT		PTs: 10043554 Thrombocytopenia 10035528 Platelet count decreased
Neutropenia	PT		PTs: 10016288 Febrile neutropenia 10029354 Neutropenia 10029366 Neutrophil count decreased
Hemorrhagic Events	SMQ	Narrow	SMQ 20000038 Hemorrhages
Serious Hemorrhagic Events Concurrent[a] with a Grade 3 or 4 Thrombocytopenia	SMQ/PT	Narrow	SAEs in the Hemorrhagic Events search concurrent with a Grade 3 or 4 AE in the Thrombocytopenia search
Serious Infections	SOC		SAEs in the SOC of Infections and Infestations
Serious Infections Concurrent[a] with Grade 3 or 4 Neutropenia	SOC/PT		SAEs in the Serious Infections search concurrent with a Grade 3 or 4 AE in the Neutropenia search
Drug-Induced Liver Injury (DILI)[b]	SMQ	Broad/ Narrow	SMQs: 20000006 Drug Related Hepatic Disorders – Comprehensive (Broad) 20000007 Drug Related Hepatic Disorders – Severe Events Only (Narrow)
Skin Cancer or Second Malignancies	SMQ	Narrow	SMQs: 20000173 Skin Neoplasms, Malignant and Unspecified 20000194 Malignant tumours

a. An AE will be considered concurrent with the cytopenia event if the onset of the event was no more than 7 days prior to the onset of the cytopenia event and no more than 7 days after the end of the cytopenia event.

b. The Drug Related Hepatic Disorders – Comprehensive SMQ (20000006) includes all PTs in the Drug Related Hepatic Disorders - Severe Events Only SMQ (20000007). Summaries will be provided for each SMQ.