Official Title: A Randomized, Double-Blind, Placebo-Controlled Study of the Combination of PI3Kδ Inhibitor Parsaclisib and Ruxolitinib in Participants With Myelofibrosis

NCT Number: NCT04551066

Document Date: Statistical Analysis Plan Amendment 1: 26-Apr-2022

Statistical Analysis Plan

INCB 50465-313

A Randomized, Double-Blind, Placebo-Controlled Study of the Combination of PI3Kδ Inhibitor Parsaclisib and Ruxolitinib in Participants With Myelofibrosis

IND Number:	**147,207**
EudraCT Number	**2020-003130-21**
Sponsor:	**Incyte Corporation** **1801 Augustine Cut-Off** **Wilmington, DE 19803** **United States**
Protocol Version:	**Amendment 1 dated 11 OCT 2021**
CRF Approval Date:	**12 NOV 2020**
SAP Version:	**Amendment 1**
SAP Author:	[redacted] [redacted], **Biostatistics**
Date of Plan:	**26 APR 2022**

This study is being conducted in compliance with Good Clinical Practice, including the archiving of essential documents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
AE	adverse event
AESI	adverse events of special interest
ALT	alanine aminotransferase
AST	aspartate aminotransferase
CI	confidence interval
CMH	Cochran Mantel-Haenszel
CRF	case report form
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DIPSS	Dynamic International Prognostic Scoring System
DMC	Data Monitoring Committee
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
[redacted]	[redacted]
HGB	hemoglobin
HSD	Hwang-Shih-DeCani
ITT	intent-to-treat
[redacted]	[redacted]
JAK	Janus kinase
KM	Kaplan-Meier
MedDRA	Medical Dictionary for Regulatory Activities
MF	myelofibrosis
MFSAF	Myelofibrosis Symptom Assessment Form
MRI	magnetic resonance imaging
NCI	National Cancer Institute
OS	overall survival
PET-MF	post–essential thrombosis myelofibrosis
[redacted]	[redacted]
PI3K	phosphoinositide 3-kinase
PP	per protocol
PPV-MF	post–polycythemia vera myelofibrosis
[redacted]	[redacted]

Abbreviation	Term
PT	preferred term
QD	once daily
[redacted]	[redacted]
RBC	red blood cell
SAP	Statistical Analysis Plan
SOC	system organ class
TEAE	treatment-emergent adverse event
TSS	total symptom score
ULN	upper limit of normal
WHO	World Health Organization
[redacted]	[redacted]

1. INTRODUCTION

This is a Phase 3, randomized, double-blind study of the combination of the PI3Kδ inhibitor parsaclisib or matching placebo and the JAK 1/2 inhibitor ruxolitinib in participants with PMF or secondary MF (PPV-MF or PET-MF) with DIPSS risk category of intermediate or high. Prospective participants must have not received prior MF therapy with a JAK inhibitor or a PI3K inhibitor. After participants have been determined to be eligible for the study and completed the baseline symptom diary assessment for 7 days, they will be randomized to 1 of 2 treatment groups, with stratification for baseline platelet count ($\geq 100 \times 10^9$/L vs 50 to $< 100 \times 10^9$/L inclusive) and DIPSS risk category (high vs intermediate-2 vs intermediate-1). If a participant's platelet count has decreased to $< 50 \times 10^9$/L at baseline, the platelet count at screening will be used for stratification/randomization. Participants will receive study drug or matching placebo at a dose of 5 mg QD beginning on Day 1 and continuing for the duration of their participation in the study. Ruxolitinib dosing will also begin on Day 1, and dose level and protocol-allowed dose increases will be determined by baseline platelet count. Section 2 of the Protocol provides a detailed description of the investigational product, target patient population, rationale for doses to be examined, and potential risks and benefits of treatment with parsaclisib and ruxolitinib.

The purpose of this SAP is to provide details of the statistical analyses that have been outlined in the INCB 50465-313 Protocol.

2. STUDY INFORMATION, OBJECTIVES, AND ENDPOINTS

2.1. Protocol and Case Report Form Version

This SAP is based on INCB 50465-313 Protocol Amendment 1 (dated 11 OCT 2021) and CRF dated 12 NOV 2020. Unless superseded by an amendment, this SAP will be effective for all subsequent Protocol amendments and eCRF versions.

2.2. Study Objectives and Endpoints

Table 1 presents the objectives and endpoints.

Table 1: Objectives and Endpoints

Objectives	Endpoints
Primary	
To evaluate and compare the efficacy of parsaclisib plus ruxolitinib versus placebo plus ruxolitinib on spleen volume at Week 24.	Proportion of participants achieving ≥ 35% reduction in spleen volume from baseline to Week 24 as measured by MRI (or CT scan in applicable participants).
Secondary	
To evaluate and compare the effect of parsaclisib plus ruxolitinib versus placebo plus ruxolitinib on participant reports of MF symptoms.	Proportion of participants who have a ≥ 50% reduction in TSS from baseline to Week 24 as measured by the MFSAF v4.0 diary. Change in TSS from baseline to Week 24 as measured by the MFSAF v4.0 diary. Time to the first ≥ 50% reduction in TSS as measured by the MFSAF v4.0 diary.
To evaluate and compare the effect of parsaclisib plus ruxolitinib versus placebo plus ruxolitinib with respect to OS.	OS determined from the date of randomization until death due to any cause.
To evaluate and compare the safety and tolerability of parsaclisib plus ruxolitinib versus placebo plus ruxolitinib.	Safety and tolerability will be assessed by monitoring the frequency and severity of AEs, performing physical examinations, and evaluating changes in vital signs, ECGs, and laboratory results.
To evaluate and compare the time to onset or response and duration of response in spleen volume of participants receiving parsaclisib plus ruxolitinib versus placebo plus ruxolitinib.	Time of onset of a ≥ 35% reduction in spleen volume and duration of maintenance of a ≥ 35% reduction in spleen volume.
[redacted]	

Table 1: Objectives and Endpoints (Continued)

Objectives	Endpoints
[redacted]	

3. STUDY DESIGN

3.1. Randomization

Participants will be randomized 1:1 to parsaclisib plus ruxolitinib and placebo plus ruxolitinib with stratification for DIPSS risk category (intermediate-1 vs intermediate-2 vs high) and baseline platelets ($\geq 100 \times 10^9$/L vs 50 to $< 100 \times 10^9$/L inclusive). Note that if a participant's platelet count is $< 50 \times 10^9$/L at baseline, the platelet count at screening will be used for stratification/randomization.

3.2. Control of Type I Error

This study defines a single primary endpoint with a single primary analysis. The study will be claimed to have achieved the primary efficacy objective when the primary endpoint shows a significant result at 2-sided alpha of 0.05 at final analysis. The secondary efficacy endpoints will be analyzed only when the study has reached the efficacy objective in the primary endpoint. The key secondary efficacy variables will be tested following a fixed-sequence-testing procedure with each at the 2-sided alpha level of 0.05 in the order below:

1. The proportion of participants who have a ≥ 50% reduction from baseline to Week 24 in MFSAF TSS
2. Overall survival

Other secondary [redacted] endpoints will be tested using a 2-sided, 5% significance level; however, these are not considered alpha-controlled hypotheses.

There are 2 planned interim analyses in this study, both aiming to terminate the study if there is inadequate efficacy. The nonbinding futility rules in Interim Analyses 1 and 2 will control the probability of continuing beyond Interim Analysis 2 at 10%. As no formal decision for efficacy will be made based on these interim analyses, no overall alpha will be spent.

An independent DMC will be assembled to monitor safety data, efficacy data, and study conduct on a regular and ongoing basis during the study. See Section 10 for details regarding the alpha-spending plan and interim analyses conducted in this study.

3.3. Sample Size Considerations

The primary endpoint of this study is the proportion of participants with ≥ 35% spleen volume reduction from baseline to Week 24 as measured by MRI (or CT scan in applicable participants). The primary endpoint will be analyzed by CMH test stratified by DIPSS category (intermediate-1 vs intermediate-2 vs high) and baseline platelet count ($\geq 100 \times 10^9$/L vs 50 to $< 100 \times 10^9$/L inclusive). Assuming a 58.2% (29.4%) response rate for participants with normal (low) platelet count for the ruxolitinib plus parsaclisib group versus 41.0% (17.2%) response rate for participants with normal (low) platelet count for the ruxolitinib plus placebo group, to achieve 83.67% power to detect a statistically significant difference with a 2-sided Type I error of 5%, a sample size of 440 participants is required to be randomized equally between treatment groups.

For the key secondary endpoint of proportion of participants who have a ≥ 50% in MFSAF TSS from baseline to Week 24, assuming a response rate of 61.7% (50.0%) for participants with

normal (low) platelet count for the ruxolitinib plus parsaclisib group and a response rate of 45.9% (34.5%) for participants with normal (low) platelet count for the ruxolitinib plus placebo group, a sample size of 440 will provide 80.02% power using CMH test given a 2-sided Type I error of 5%.

A test for OS will be performed when 85 deaths are observed using the log-rank test stratified by platelet count and DIPSS category at baseline at a 2-sided Type I error level of 0.05. The log-rank test will provide 80% power to detect a hazard ratio of 0.54. Assuming a median survival time of 5.3 years for ruxolitinib plus placebo and enrollment duration is 1.5 years, the expected time to analysis from the first participant randomized is 3 years.

3.4. Schedule of Assessments

Refer to the Protocol for a full description of all study procedures and assessment schedules for this study.

4. DATA HANDLING DEFINITIONS AND CONVENTIONS

4.1. Scheduled Study Evaluations and Study Periods

4.1.1. Day 1

Day 1 is the date that the first dose of parsaclisib, placebo, or ruxolitinib is administered to the participants.

For randomized participants not treated with parsaclisib, placebo, or ruxolitinib, Day 1 is defined as the date of randomization.

4.1.2. Day 1 of Crossover for Crossover Participants

For participants who cross over from placebo plus ruxolitinib to parsaclisib plus ruxolitinib, Day 1 of crossover is the date the first dose of open-label parsaclisib is administered to the participants.

4.1.3. Study Day

If a visit/reporting date is on or after Day 1, then the study day at the visit/reporting date will be calculated as

Day # = (visit/reporting date – Day 1 date + 1).

If the visit/reporting date is before Day 1, then the study day at the visit/reporting date will be calculated as

Day # = (visit/reporting date – Day 1 date).

A study day of –1 indicates 1 day before Day 1.

4.1.4. Baseline Value

Baseline is the last nonmissing measurement obtained before the first administration of parsaclisib, placebo, or ruxolitinib.

For randomized participants not treated with any study drug, baseline is defined as the last nonmissing measurement obtained before the date of randomization for all parameters.

When scheduled assessments and unscheduled assessments occur on the same day and the time of the assessment or time of first dose is not available, use the following convention to determine baseline:

- If both a scheduled and an unscheduled visit are available on the day of the first dose and the time is missing, use the scheduled assessment as baseline.
- If all scheduled assessments are missing on the day of the first dose and an unscheduled assessment is available, use the unscheduled assessment as baseline.

4.1.5. Handling of Missing and Incomplete Dates

In general, values for missing dates will not be handled unless methods for handling missing dates are specified in this section or relevant sections. The original reported dates collected on the eCRF should be used in all relevant listings. The following rules will be used for handling partial dates for analyses requiring dates.

When calculating the time since diagnosis of disease, a partial disease diagnosis date will be handled as follows in the calculation:

- If only the day is missing, then the first day of the month will be used.
- If both the month and day are missing, then 01 JAN of the year will be used.
- If the diagnosis date is completely missing, then the time since diagnosis will not be calculated.

When the date of the last dose is used in deriving variables such as duration of treatment or TEAE flag, a missing or partial date of the last dose will be handled as follows:

- If only the day is missing, then the earlier date of the last day of the month or the date that the participant discontinued treatment will be used.
- Otherwise, the date that the participant discontinued treatment will be used as the date of the last dose.

For relevant efficacy endpoints, a partial date of the death date will be handled as follows in the calculation:

- If mmyyyy for the last known alive date = mmyyyy for the death date, then the death date will be set to the day after the last known alive date.
- If mmyyyy for the last known alive date < mmyyyy for the death date, then the death date will be set to the first day of the death month.
- Otherwise, the partial death date will not be imputed.

4.2. Variable Definitions

The following variables will only be calculated if not reported on the eCRF.

4.2.1. Body Mass Index

Body mass index will be calculated as follows:

Body mass index (kg/m^2) = [weight (kg)] / [height (m)]^2.

4.2.2. Prior and Concomitant Medication

Prior medication is defined as any nonstudy medication started before the first dose of parsaclisib, placebo, or ruxolitinib.

Concomitant medication is defined as any nonstudy medication that is started accordingly:

- Before the date of first administration of parsaclisib, placebo, or ruxolitinib and is ongoing throughout the study or ends on/after the date of first study drug administration.
- On/after the date of first administration of parsaclisib, placebo, or ruxolitinib and is ongoing or ends during the course of study.

A prior medication could also be classified as "both prior and concomitant medication" if the end date is on or after first dose of parsaclisib, placebo, or ruxolitinib. In the listing, it will be indicated whether a medication is only prior, only concomitant, or both prior and concomitant.

For the purposes of analysis, all medications will be considered concomitant medications unless the medications can unequivocally be defined as not concomitant.

5. STATISTICAL METHODOLOGY

5.1. General Methodology

Unless otherwise noted, SAS® software (SAS Institute Inc, Cary, NC; v9 or later) will be used for the generation of all tables, graphs, and statistical analyses. Descriptive summaries for continuous variables will include but not be limited to the number of observations, mean, standard deviation, median, minimum, and maximum. Descriptive summaries for categorical variables will include the number and percentage of participants in each category.

A participant randomized to the placebo group may cross over to parsaclisib treatment if the participant meets the criteria for an early crossover defined in Protocol Section 4.1.3. Unless specified, efficacy and safety analyses for between-treatment comparisons (parsaclisib vs placebo) will be conducted using the data truncated at the time of treatment crossover for those participants who make an early treatment crossover.

Interim analyses are planned for this study as defined in Section 10.

5.2. Treatment Groups

This is a randomized, double-blind, parallel treatment group design. Participants will be summarized by treatment groups. A participant randomized to the control group may make a crossover to the active treatment before Week 24 if the participant met the criteria for an early crossover. Unless specified, efficacy and safety analyses for between-treatment comparisons (parsaclisib + ruxolitinib vs placebo + ruxolitinib) will be conducted using the data truncated at the time of treatment crossover for those participants who made an early treatment crossover before Week 24. Data collected after the treatment crossover will be summarized and tabulated separately without treatment comparisons.

5.3. Analysis Populations

5.3.1. Intent-to-Treat Population

All participants who are randomized will constitute the ITT population. Treatment groups for this population will be defined according to the treatment assignment at the time of randomization regardless of the actual study drug the participant might take during their participation in the study.

The ITT population will be used for the summary of demographics, baseline characteristics, participant disposition, and analyses of all efficacy data.

Participants with missing baseline assessments will be excluded from the ITT population for the responder analyses.

5.3.2. Per-Protocol Population

Participants in the ITT who are considered to be sufficiently compliant with the Protocol will compose the PP population.

The following procedures will be performed to identify those participants who are to be excluded from the PP population:

- Clinical review of Protocol deviations
- Clinical review of concomitant medications as defined in Section 6.8 of the Protocol
- Clinical review of the dose administration and drug accountability listing
- Clinical review of inclusion/exclusion criteria

Participants with missing baseline assessments will be excluded from the PP population for the responder analyses.

The determination of participants being considered for exclusion from the PP population by the clinical team will be prepared and signed before unblinding and database freeze.

The PP population will be used in the supportive sensitivity analyses for efficacy endpoints.

5.3.3. Safety Population

The safety population will include all participants who received at least 1 dose of parsaclisib, placebo, or ruxolitinib. Treatment groups for this population will be determined according to the actual treatment the participant received regardless of assigned study drug treatment.

All safety analyses will be conducted using the safety population.

6. BASELINE, EXPOSURE, AND DISPOSITION

Appendix A provides a list of data displays. Sample data displays are included in a separate document.

6.1. Demographics, Baseline Characteristics, and Disease History

6.1.1. Demographics and Baseline Characteristics

The following demographics and baseline characteristics will be summarized for the ITT population by treatment group: age, sex, race, ethnicity, weight, height, and BMI.

6.1.2. Baseline Disease Characteristics

The following baseline disease characteristics will be summarized for the ITT population by treatment group: ECOG performance status, neutrophils, platelets, hemoglobin, and leukocytes at baseline.

6.1.3. Disease History

6.1.3.1. Myelofibrosis Disease History

The time since diagnosis to randomization, current MF disease, prognostic factors, risk level as defined by DIPSS, a prior splenic irradiation was performed, if splenectomy was performed, and whether transfusion was received in the 12 weeks before randomization will be summarized for all participants in the ITT population by treatment group.

Time since diagnosis will be calculated as follows:

Time since diagnosis (years) = (date of randomization – date of diagnosis + 1) / 365.25.

The MF screening symptom individual and total scores will be summarized for the ITT population by the treatment group.

6.1.4. Prior Therapy

The number of prior systemic regimens for MF will be summarized for all participants in the ITT population by treatment group. The component drugs of prior systemic therapy regimens will be coded using the WHO Drug Dictionary. The number and percentage of participants who received each drug will be summarized by WHO drug class and WHO drug PT. The regimen name, route, dose, frequency, start and stop dates, and reason for discontinuation will be listed.

The number of participants who received prior radiation for non-MF related cancer treatment will be summarized for the ITT population by treatment group. The radiotherapy type, body site, start and stop dates, reason for regimen, total dose, number of fractions received, and best response will be listed.

The number of participants who had prior surgery or surgical procedure for non-MF related cancer treatment will be summarized for the ITT population by treatment group. The date and description of the surgery/procedure will be listed.

6.1.5. Medical History

For participants in the ITT population, medical history will be summarized by assigned treatment group. This summation will include the number and percentage of participants with significant medical history for each body system/organ class as documented on the eCRF.

6.2. Disposition of Participant

The number and percentage of participants who were randomized, who were treated, who were ongoing with study treatment, who discontinued study treatment with a primary reason for discontinuation, who were still in the study, and who withdrew from the study with a primary reason for withdrawal will be summarized for the ITT population by treatment group. The number of participants randomized by country and/or site will also be provided by treatment group.

6.3. Protocol Deviations

Protocol deviations recorded in the clinical trial management system will be summarized and listed.

6.4. Exposure

For participants in the safety population, exposure to parsaclisib, placebo, and ruxolitinib will be summarized descriptively as the following:

- **Duration of treatment with ruxolitinib (days):** date of last dose of ruxolitinib – date of first dose of ruxolitinib + 1.

- **Duration of treatment with parsaclisib/placebo (days):** date of last dose of parsaclisib/placebo – date of first dose of parsaclisib/placebo + 1.
- **Average daily dose of ruxolitinib (mg/day):** total actual ruxolitinib dose taken (mg) / duration of treatment with ruxolitinib (days).
- **Average daily dose of parsaclisib/placebo (mg/day):** total actual parsaclisib/placebo dose taken (mg) / duration of treatment with parsaclisib/placebo (days).

  Total actual dose taken will be calculated based on the information entered on the Drug Exposure eCRF.
- **Total dose administered for ruxolitinib (mg):** total dose administered for each participant will be the sum of ruxolitinib administered across cycles.
- **Total dose administered for parsaclisib/placebo (mg):** total dose administered for each participant will be the sum of parsaclisib/placebo administered across cycles.

6.5. Study Drug Compliance

For participants in the safety population, overall compliance (%) for parsaclisib and ruxolitinib will be calculated for all participants as follows:

compliance (%) = 100 × [total dose actually taken] / [total prescribed dose].

The total prescribed dose is defined as the sum of the doses prescribed by the investigator accounting for dose modifications.

The total actual dose taken will be calculated based on information entered on the Drug Accountability eCRF. If there are dispensed drugs that have not been returned yet, the actual dose taken starting from the dispense date of the unreturned drugs will be imputed by the dose taken as reported on the Dosing eCRF.

6.6. Prior and Concomitant Medication

Prior medications and concomitant medications will be coded using the WHO Drug Dictionary. The number and percentage of participants in the ITT population for each prior and concomitant medication will be summarized by WHO drug class and WHO drug PT for each treatment group.

7. EFFICACY

Appendix A provides a list of data displays. Sample data displays are included in a separate document.

7.1. General Considerations

Unless otherwise stated, the strata identified in the randomization process will be used in all efficacy analyses. Within a platelet category, in the event that the study fails to enroll a sufficient number of participants in a particular DIPSS stratification level, at the time of final analysis for database lock, that stratification will be combined with the nearest adjacent DIPSS

level (eg, intermediate-1 vs intermediate-2 vs high). For example, if either intermediate-1 or high risk is too small (eg, fewer than 10 participants), it will be merged with intermediate-2. If intermediate-2 has too few participants (eg, fewer than 10 participants), then it will be merged with the smaller of the 2 adjacent strata. At the time of interim analyses, all DIPSS stratification levels will be combined if too few participants (eg, fewer than 10 participants) are enrolled in 1 DIPSS stratification level.

The ITT population will be used for all efficacy analyses unless otherwise stated.

7.2. Efficacy Hypotheses

The primary hypothesis is that the number of participants achieving ≥ 35% reduction in spleen volume from baseline to Week 24 for parsaclisib plus ruxolitinib is not the same as the number of participants achieving ≥ 35% reduction in spleen volume from baseline to Week 24 when compared with placebo plus ruxolitinib as measured by the odds ratio for response. The hypotheses of the study are as follows:

- H_0 (null hypothesis): $\theta_{PAR} = \theta_{PLB}$
- H_A(alternative hypothesis): $\theta_{PAR} \neq \theta_{PLB}$

7.3. Analysis of the Primary Efficacy Parameter

7.3.1. Primary Efficacy Analysis

The primary endpoint of this study is the proportion of participants with ≥ 35% spleen volume reduction from baseline to Week 24 as measured by MRI (or CT scan in applicable participants). The primary endpoint will be analyzed by CMH test stratified by DIPSS category (intermediate-1 vs intermediate-2 vs high) and baseline (or screening, if applicable) platelet count (≥ 100×10^9/L vs 50 to < 100×10^9/L inclusive).

The baseline spleen volume will be measured by MRI (or by CT for applicable participants) during the baseline visit (Days −7 to 1); the last value will be used if there are multiple values measured during the baseline visit.

The Week 24 spleen volume will be measured by MRI (or by CT for applicable participants) during the Week 24 visit; the last value will be used if there are multiple values measured during the Week 24 visit. Missing values will not be imputed.

The percentage change from baseline to Week 24 will be calculated using the following formula:

% change = 100 × [Week 24 spleen volume – baseline spleen volume] / baseline spleen volume.

A participant will be considered as having achieved ≥ 35% reduction of the spleen volume from baseline to Week 24, if the participant had both baseline and Week 24 spleen volume measurements, the participant did not make an early crossover to receive the active treatment (for ruxolitinib plus placebo participants) before the Week 24 visit, the percentage change from baseline was not missing, and the percentage change from baseline reduction was ≥ 35%.

A participant who had a missing Week 24 spleen volume, discontinued treatment before Week 24, or made an early crossover to receive parsaclisib plus ruxolitinib before the Week 24 visit will be considered as a nonresponder for the endpoint of ≥ 35% reduction of spleen volume.

A participant who had a missing baseline assessment will be considered as nonevaluable.

At the time of interim analysis, a participant who was still on treatment, but has not followed up for the timepoint assessment and had no timepoint assessment, will also be considered as nonevaluable.

7.3.2. Subgroup Analyses for Primary Endpoint

Subgroups will be formed based on the following participant characteristics and baseline variables for those participants whose data are available:

- Sex: male and female
- Age group: ≤ 65 years or > 65 years
- MF type: PMF or PPV-MF/PET-MF
- Presence/absence of V617F mutation at baseline
- Presence/absence of HMR mutations at baseline
- Baseline spleen volume group: $\leq$ median or $>$ median
- Baseline spleen palpation size group: ≤ 10 cm or > 10 cm
- Race: White/Caucasian, Black/African-American, Asian, American-Indian/Alaska Native, Native Hawaiian/Pacific Islander, or other
- Ethnicity: Hispanic/Latino, not Hispanic/Latino, or other
- Platelet: $\geq 100 \times 10^9$/L vs 50 to $< 100 \times 10^9$/L, inclusive using baseline platelet counts. Note that if a participants's platelet count is $<50 \times 10^9$/L at baseline, the platelet count at screening will be used for stratification/randomization.
- DIPSS: intermediate-1, intermediate-2, or high

7.3.3. Sensitivity and Supportive Analyses for Primary Endpoint

The primary endpoint will be analyzed using the PP population as a sensitivity analysis to the ITT population.

The participants who make an early crossover and later turn into responders will be considered responders and analyzed in the placebo group as a supportive analysis.

7.4. Analysis of the Secondary Efficacy Parameters

7.4.1. Spleen Volume Reduction

7.4.1.1. Time to the First ≥ 35% Reduction in Spleen Volume

The time to the first $\geq 35\%$ reduction in spleen volume is defined as the time from randomization to the first time participants had $\geq 35\%$ reduction in spleen volume. Participants with baseline and postbaseline MRI or CT scan who do not have $\geq 35\%$ reduction in spleen volume at the time of analysis will be censored at the time of the last MRI or CT scan. If the participants have no baseline or postbaseline MRI or CT scan, they will be censored at the date of randomization.

The stratified log-rank test will be used to compare time to the first ≥ 35% reduction in spleen volume between 2 treatment groups in the ITT population at a 1-sided 2.5% significance level, stratified for baseline (or screening, if applicable) platelet count and DIPSS category. The strata identified in the randomization process will be used for the analysis.

Kaplan-Meier curves for time to the first ≥ 35% reduction in spleen volume will be presented by treatment groups. The KM estimate of median time to the first ≥ 35% reduction in spleen volume will be presented with its 95% CI. The 95% CI will be calculated using the generalization of method of Brookmeyer and Crowley (1982) with log-log transformation (Klein and Moeschberger 1997).

7.4.1.2. Duration of Maintenance of a ≥ 35% Reduction in Spleen Volume

The duration of ≥ 35% reduction from baseline in spleen volume is defined as the interval between the first spleen volume measurement that is ≥ 35% reduction from baseline and the date of the first measurement that is no longer a ≥ 35% reduction from baseline. If the end date is not observed before the database cutoff, the duration will be censored at the last assessment.

The endpoint will be derived for participants who were randomized and who had at least 1 measurement of ≥ 35% reduction from baseline.

Kaplan-Meier estimates of median duration with 95% CIs will be provided.

The ITT population with spleen volume response before crossover will be used.

A supportive analysis for duration of ≥ 35% reduction from baseline in spleen volume defined as the interval between the first spleen volume measurement that is ≥ 35% reduction from baseline and the date of the first measurement that is no longer a ≥ 35% reduction from baseline that is also a > 25% increase from nadir will be provided.

7.4.2. MFSAF Total Symptom Score v4.0

The MFSAF v4.0 is composed of 7 individual symptom scores (fatigue, night sweats, itchiness, abdominal discomfort, pain under left ribs, early satiety, bone pain), each collected daily with a 0- to 10-point scale.

The daily MFSAF TSS is the sum of the 7 individual symptom scores collected on the same day. The MFSAF TSS will be missing if there are any missing individual scores. Observations with missing dates will be excluded from the analysis.

The baseline total score is defined as the average of daily total scores from the last 7 days before the first dose of parsaclisib, placebo, or ruxolitinib; the baseline total score will be missing if there are ≥ 4 missing out of the 7 daily total scores.

The Week 12 total score will be the average of daily total scores from the first 7 consecutive days starting Day 78; the Week 12 total score will be missing if there are ≥ 4 missing out of the 7 daily total scores.

The Week 24 total score will be the average of daily total scores from the 7 consecutive days before the Week 24 visit; the Week 24 total score will be missing if there are ≥ 4 missing out of the 7 daily total scores.

The percentage change from baseline to Week 12/Week 24 will be calculated using the following formula:

% change = 100 × [Week 12/Week 24 MFSAF TSS – baseline MFSAF TSS] / baseline MFSAF TSS.

7.4.2.1. Windowing of MFSAF v4.0 Total Symptom Score

7.4.2.1.1. MFSAF v4.0 Total Symptom Score for Non – Time-to-Event Endpoints

The MFSAF v4.0 TSS for baseline, Week 12, and Week 24 will be determined by averaging the daily MFSAF v4.0 TSS for the days between the start and end of windows as described in Table 2. By-question summaries for the 7 individual symptom scores that compose the MFSAF TSS at baseline, Week 12, and Week 24 will use the same windowing algorithm as used for the daily TSS.

The TSS and the 7 individual scores at Weeks 4, 8, 16, and 20 will be derived in a similar fashion using the first nonmissing daily TSS available between

Day (7 × week) − 6

and

Day (7 × week)

as the start of the window.

Table 2: Window for Deriving MFSAF v4.0 Total Symptom Score for Baseline, Week 12, and Week 24

Period	Start of Window	End of Window	Missing
Baseline	7 days on or before end of window	Last day that a daily TSS was collected between Day −7 and Day −1 (inclusive)	4 or more missing of the 7 daily TSSs in the window
Week 12	First day a daily TSS was collected between Day 78 and Day 84 (inclusive)	7 days on or after start of window	4 or more missing of the 7 daily TSSs in the window
Week 24	7 days on or before end of window	The day before Week 24 visit	4 or more missing of the 7 daily TSSs in the window

7.4.2.1.2. MFSAF v4.0 Total Symptom Score for Time-to-Event Endpoints

The MFSAF v4.0 TSS for baseline will be derived as the same way in Section 7.4.2.1.1. The daily postbaseline TSS will be calculated from Day 7 to the day before the Week 24 visit. The Day X TSS is the average TSS between Day (X – 6) and Day X. For example, the TSS on Day 7 is the average of the TSS from Day 1 to Day 7. The missing individual score and the missing daily TSS will follow the rule defined in Section 7.4.2.

7.4.2.2. Proportion of Participants With a ≥ 50% Reduction in MFSAF v4.0 Total Symptom Score

The difference between parsaclisib plus ruxolitinib and placebo plus ruxolitinib in the proportion of participants who have a ≥ 50% reduction in TSS from baseline to Week 24 evaluated by MFSAF v4.0 will be assessed using CMH test stratified by baseline (or screening, if applicable) platelet count and DIPSS category at randomization for the ITT population as primary analysis and PP population as sensitivity analysis. In the case that there is an insufficient number of participants enrolled in any DIPSS level within a platelet category, the collapsing rule specified in Section 7.3.1 will be applied. Missing values will not be imputed.

The percentage change from baseline to Week 24 will be calculated using the following formula:

% change = 100 × (Week 24 total score – baseline total score) / baseline total score.

A participant will be considered as having achieved ≥ 50% reduction of the TSS from baseline to Week 24 if the participant had both baseline and Week 24 TSS, the participant did not make an early crossover to receive parsaclisib plus ruxolitinib before the Week 24 visit, and the percentage change from baseline was not missing and was ≤ −50%.

A participant who has a missing Week 24 TSS, discontinued treatment before Week 24, is unblinded prior to Week 24 for early crossover to parsaclisib plus ruxolitinib, or has a baseline value of 0 will be considered as having not achieved the ≥ 50% reduction.

A participant who has a missing baseline assessment will be considered as nonevaluable.

At the time of interim analysis, a participant who is still on treatment, but has no timepoint assessment for TSS, will also be considered as nonevaluable.

7.4.2.3. Mean Change and Percentage Change in MFSAF v4.0 Total Symptom Score

The mean change and percentage change from baseline to Week X will be calculated using the following formula:

% change = 100 × (Week X total score – baseline total score) / baseline total score.

The mean change and percentage change in TSS, the 7 individual scores, the total score of the 3 spleen-related symptom, including abdominal discomfort, pain under left ribs, and early satiety from baseline to Week X, will be summarized descriptively by treatment group and every 4 weeks using the ITT population. Missing values will not be imputed.

A plot of percentage change from baseline in MFSAF TSS at each visit will be provided.

7.4.2.4. Time to the First ≥ 50% Reduction in Myelofibrosis Symptom Assessment Form v4.0 Total Symptom Score

The time to the first ≥ 50% reduction in TSS is defined as the time from randomization to the first time participants had ≥ 50% reduction in TSS. The daily TSS will be calculated according to the algorithm defined in Section 7.4.2.1.2. Participants with valid calculated baseline TSS and at least 1 postbaseline TSS who do not have ≥ 50% reduction in TSS at the time of analysis will be censored at the time of the last valid calculated TSS. If the participants have no valid calculated baseline or postbaseline TSS, they will be censored at the date of randomization.

The stratified log-rank test will be used to compare time to the first ≥ 50% reduction in TSS between 2 treatment groups in the ITT population at a 1-sided 2.5% significance level, stratified for baseline (or screening, if applicable) platelet count and DIPSS category. The strata identified in the randomization process will be used for the analysis.

Kaplan-Meier curves for time to the first ≥ 50% reduction in TSS will be presented by treatment groups. The KM estimate of median time to the first ≥ 50% reduction in TSS will be presented with its 95% CI. The 95% CI will be calculated using the generalization of method of Brookmeyer and Crowley (1982) with log-log transformation (Klein and Moeschberger 1997).

7.4.3. Overall Survival

Overall survival is defined as the interval between the randomization date and the date of death due to any cause. Date of death will be determined using the Death Report, the Survival Follow-Up, and the Participant Status eCRFs. Participants who are lost-to-follow-up or still alive at the time of analysis will be right-censored at the earlier of the date the participant was last known alive and the clinical data cutoff date for the analysis. The last known alive date is defined as the later of the last study visit and the date the participant was last known alive from the Survival Follow-Up and Participant Status eCRFs. Partial death dates will be handled using the rules described in Section 4.1.5.

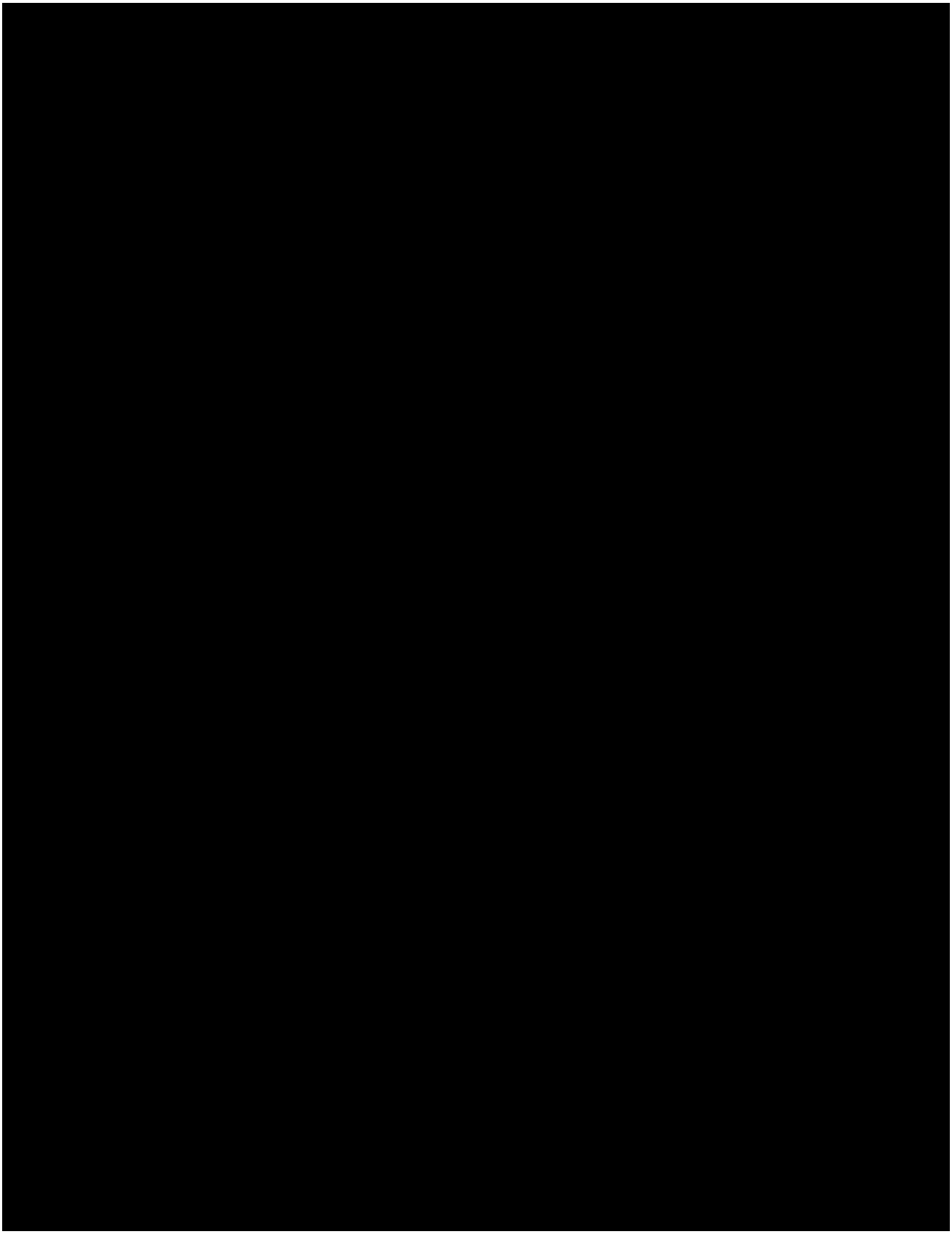
The treatment difference in survival for parsaclisib plus ruxolitinib and placebo plus ruxolitinib will be assessed by a log-rank test on the ITT population and PP population stratified for baseline (or screening, if applicable) platelet and DIPSS category. The strata identified in the randomization process will be used for the analysis. In the event that stratification level combinations have limited numbers of participants, 1 or more of the strata may be combined to better facilitate estimation of the overall HR.

The hazard ratio estimate and CIs will be estimated based on a stratified Cox proportional hazards model using the same stratification factors as for the log-rank test with Efron's method (1977) accounting for ties in death days.

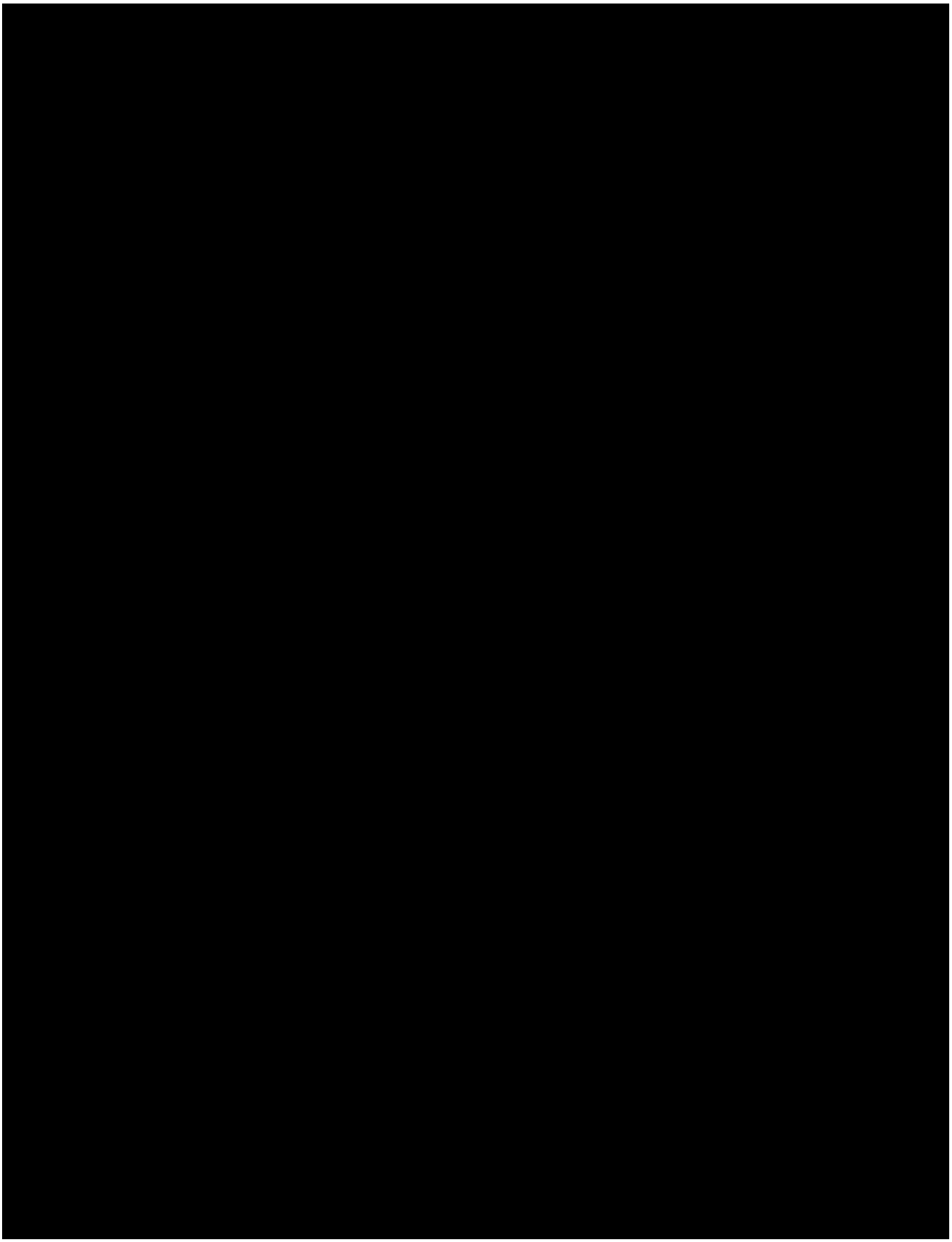
Kaplan-Meier time-to-event curves will be presented by treatment groups. Median survival will be estimated using the KM method. Confidence intervals for median survival time will be calculated using the generalization of Brookmeyer and Crowley's method (1982) with log-log transformation (Klein and Moeschberger 1997).

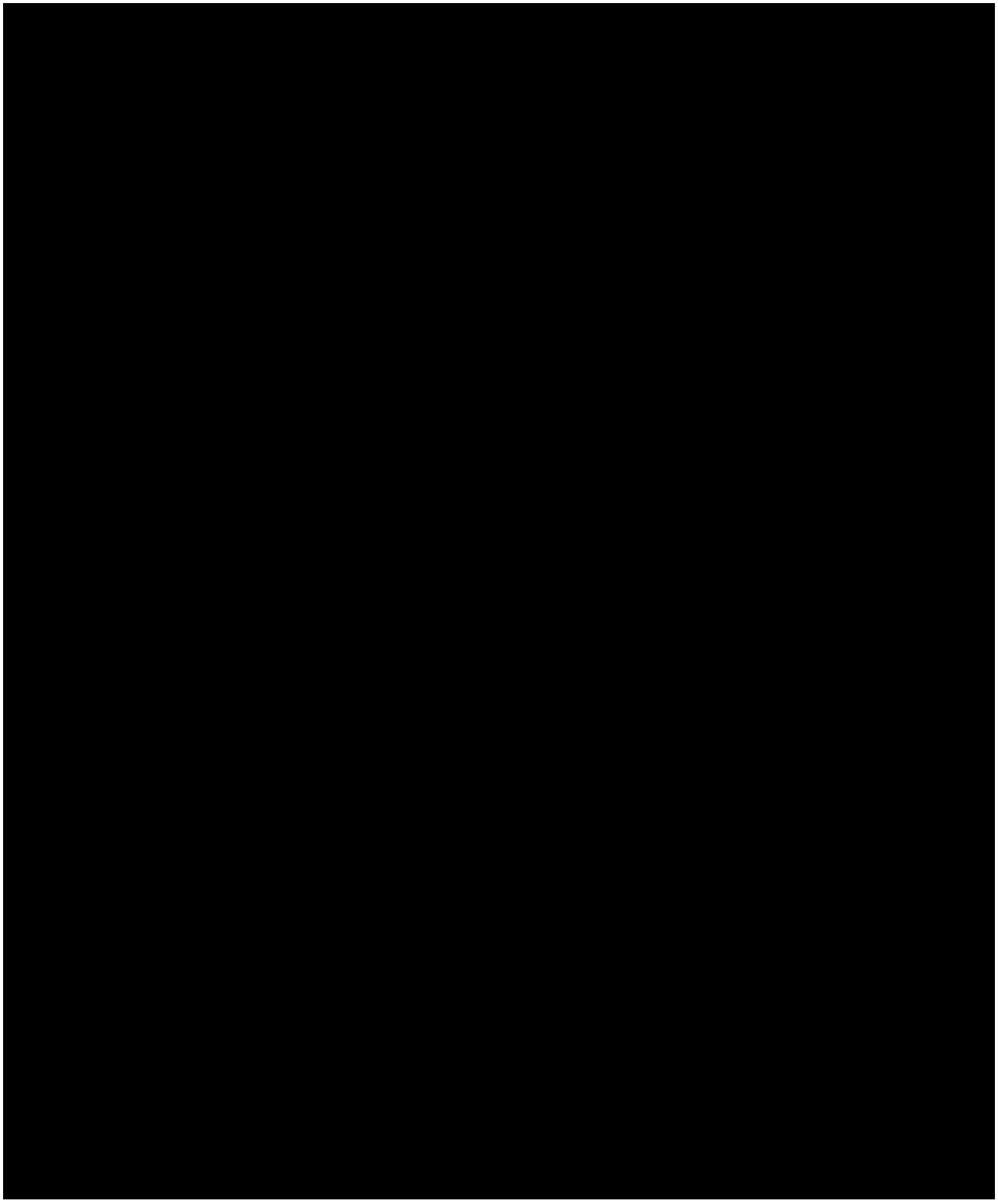
The ITT population will be used for the analysis of OS.

A sensitivity analysis of OS will be analyzed using rank preserving structural failure time method to adjust the bias in OS due to crossover.

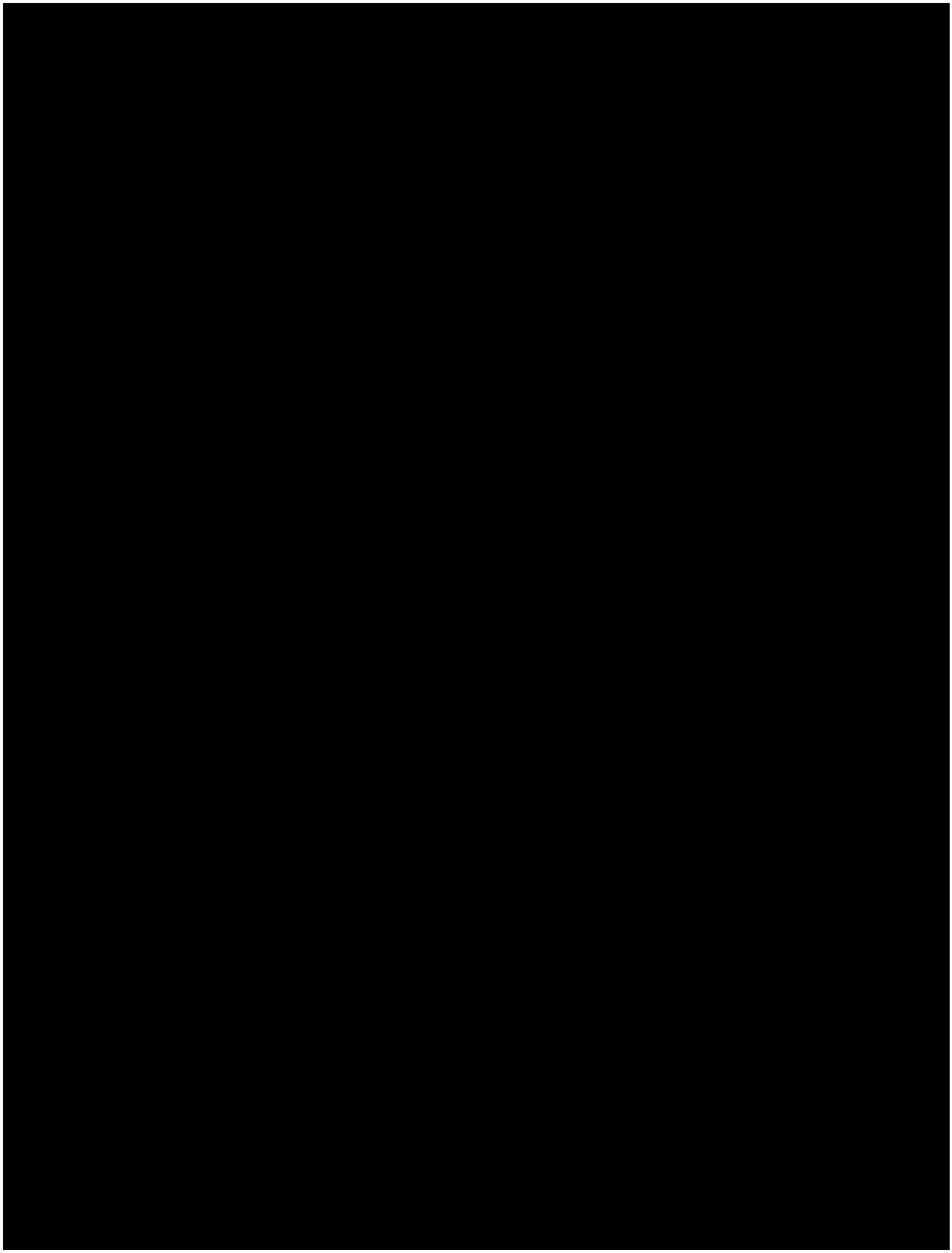


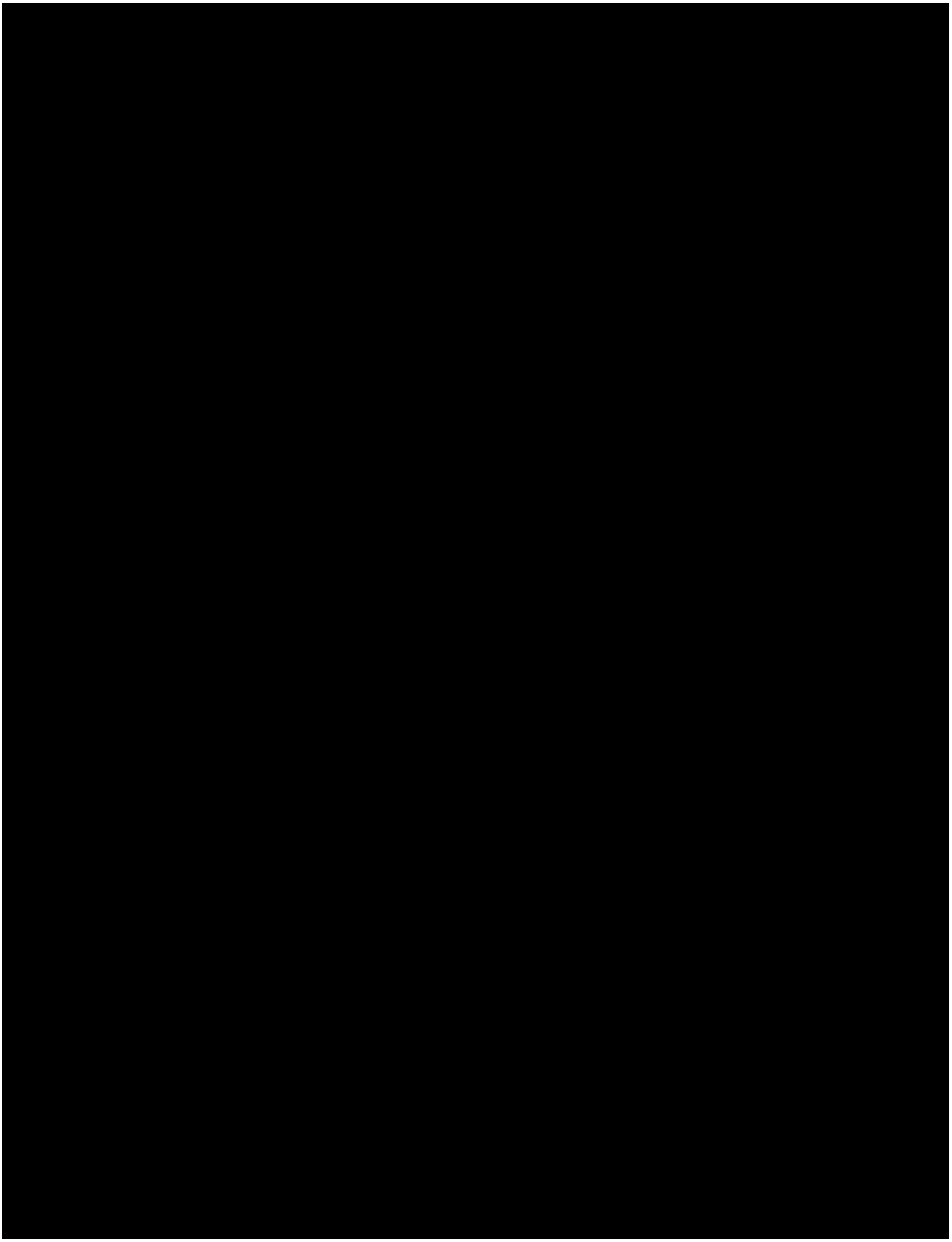
[redacted]





[redacted]





[redacted]

9. SAFETY AND TOLERABILITY

Appendix A provides a list of data displays. Sample data displays are included in a separate document.

9.1. General Considerations

Summary tables may be replaced with listings when appropriate. For instance, an AE frequency table may be replaced with a listing if it only contains a few unique PTs reported on relatively few participants.

Safety analyses for treatment comparisons will exclude data collected after treatment crossover. Data post-treatment crossover will be analyzed separately and descriptively without any treatment comparisons. Day 1 of crossover will be used for the crossover participants in the crossover phase.

9.2. Adverse Events

9.2.1. Adverse Event Definitions

A TEAE is any AE either reported for the first time or worsening of a pre-existing event after the first dose of study drug until 30 to 35 days after the last dose of parsaclisib/matching placebo or ruxolitinib. Analysis of AEs (as discussed below) will be limited to TEAEs, but data listings will include all AEs regardless of their timing in relation to study drug administration.

Adverse events will be tabulated by MedDRA PT and SOC. Severity of AEs will be graded using the NCI CTCAE v5. The CTCAE reporting guidelines and grading details are available on the Cancer Therapy Evaluation Program website.

The subset of AEs considered by the investigator to be related to study drug will be considered to be treatment-related AEs. If the investigator does not specify the relationship of the AE to study drug, the AE will be considered to be treatment-related. The incidence of AEs and treatment-related AEs will be tabulated. In addition, serious TEAEs will also be tabulated.

To summarize TEAEs using the life-table method, the time to first occurrence of a TEAE is defined as the interval between the date of first dose of study drug and the date of the first

occurrence of a TEAE. Participants who have not experienced a TEAE at the time of analysis will be right-censored at the earlier of 30 to 35 days after the last dose of study drug and the date of last study visit through the end of safety follow-up. The effective sample size is defined as the number of participants at the beginning of interval minus half the participants censored in the interval. The conditional proportion is calculated as the number of first events divided by the effective sample size in the interval. If participants have missing or partial last dose dates of study drug, the partial or missing dates will be handled using the rules explained in Section 4.1.5. Any missing onset date of a TEAE will be handled according the following rules:

- If completely missing, then Day 1 will be used.
- If only the day is missing, then the first day of the month or Day 1, whichever is later, will be used.
- If both the month and day are missing, then 01 JAN of the year or Day 1, whichever is later, will be used.

9.2.2. Adverse Events of Special Interest

The following TEAEs, regardless of seriousness, will be tabulated by MedDRA PT for the safety population for each treatment group.

- Colitis
- Diarrhea of Grade 2 or higher
- Rash of Grade 2 or higher
- Pneumonitis
- Dermatitis exfoliative
- Intestinal perforation
- Cytomegalovirus infection
- Herpes simplex infection
- Herpes (varicella) zoster virus infection
- Pneumocystis jirovecii infection
- Alanine aminotransferase increased ≥ 5 × ULN
- Aspartate aminotransferase increased ≥ 5 × ULN

Adverse events of special interest should be captured in the eCRF.

Time to the onset of the first occurrence of each AESI will be analyzed using KM method. Median time to onset of the first occurrence of each AESI will be estimated by KM. Participants with no AESI at the time of analysis will be censored at the time of analysis.

The longest duration of AESI will be analyzed using KM method for participants in the safety population with at least one occurrence of the AESI. Median longest duration of each AESI will be estimated by KM. Participants whose AESI was not resolved at the time of analysis will be

censored at the time of the analysis. Participants with non-SAEs not resolved by 35 days after the end-of-treatment visit will be censored at 35 days after the end-of-treatment visit.

The 2 consecutive AESIs with different toxicity grades, but with the start date of the later AESI overlapped with the end date of the previous AESI, are considered 1 AESI when calculating the duration of AESI.

9.2.3. Adverse Event Summaries

An overall summary of AEs by treatment group will include the following:

- Number (%) of participants who had any TEAEs
- Number (%) of participants who had any serious TEAEs
- Number (%) of participants who had any Grade 3 or higher TEAEs
- Number (%) of participants who had any TEAEs related to parsaclisib
- Number (%) of participants who had any TEAEs related to ruxolitinib
- Number (%) of participants who temporarily interrupted parsaclisib because of TEAEs
- Number (%) of participants who temporarily interrupted ruxolitinib because of TEAEs
- Number (%) of participants who permanently discontinued parsaclisib because of TEAEs
- Number (%) of participants who permanently discontinued ruxolitinib because of TEAEs
- Number (%) of participants who had parsaclisib dose reductions because of TEAEs
- Number (%) of participants who had ruxolitinib dose reductions because of TEAEs
- Number (%) of participants who had any fatal TEAEs

The following summaries will be produced by MedDRA term (if 10 or fewer participants appear in a table, a listing may be appropriate):

- Summary of TEAEs by MedDRA SOC and PT
- Summary of TEAEs by MedDRA PT in decreasing order of frequency
- Summary of TEAEs by MedDRA SOC, PT, and CTCAE grade category
- Summary of Grade 3 or higher TEAEs by MedDRA SOC and PT
- Summary of Grade 3 or higher TEAEs by MedDRA PT in decreasing order of frequency
- Summary of serious TEAEs by MedDRA SOC and PT
- Summary of serious TEAEs by MedDRA PT in decreasing order of frequency
- Summary of parsaclisib treatment-related TEAEs by MedDRA SOC and PT

- Summary of ruxolitinib treatment-related TEAEs by MedDRA SOC and PT
- Summary of parsaclisib treatment-related TEAEs by MedDRA PT in decreasing order of frequency
- Summary of ruxolitinib treatment-related TEAEs by MedDRA PT in decreasing order of frequency
- Summary of parsaclisib treatment-related TEAEs by MedDRA SOC, PT, and CTCAE grade category
- Summary of ruxolitinib treatment-related TEAEs by MedDRA SOC, PT, and CTCAE grade category
- Summary of Grade 3 or higher parsaclisib treatment-related TEAEs by MedDRA SOC and PT
- Summary of Grade 3 or higher ruxolitinib treatment-related TEAEs by MedDRA SOC and PT
- Summary of parsaclisib treatment-related serious TEAEs by MedDRA SOC and PT
- Summary of ruxolitinib treatment-related serious TEAEs by MedDRA SOC and PT
- Summary of TEAEs with a fatal outcome by MedDRA SOC and PT
- Summary of TEAEs leading to parsaclisib dose reduction by MedDRA SOC and PT
- Summary of TEAEs leading to ruxolitinib dose reduction by MedDRA SOC and PT
- Summary of TEAEs leading to parsaclisib dose interruption by MedDRA SOC and PT
- Summary of TEAEs leading to ruxolitinib dose interruption by MedDRA SOC and PT
- Summary of TEAEs leading to discontinuation of parsaclisib by MedDRA SOC and PT
- Summary of TEAEs leading to discontinuation of ruxolitinib by MedDRA SOC and PT
- Summary of TEAEs by MedDRA PT: life-table method
- Summary of TEAEs by MedDRA SOC and PT: life-table method
- Summary of treatment-emergent AESIs by MedDRA PT
- Median time to the first onset of treatment-emergent AESIs by MedDRA PT
- Median longest duration of treatment-emergent AESIs by MedDRA PT

9.3. Clinical Laboratory Tests

9.3.1. Laboratory Value Definitions

Laboratory values and change from baseline values will be summarized descriptively by visit. Baseline will be determined according to Section 4.1.4. If there are multiple values that meet the criteria for baseline, additional rules may be provided after consultation with the medical monitor to delineate which value will be defined as baseline.

Laboratory test values will be assessed for severity based on the numerical component of CTCAE v5.

9.3.2. Laboratory Value Summaries

All test results and associated normal ranges from central laboratories will be reported in SI units. Any laboratory test results and associated normal ranges from local laboratories will be converted to SI units.

When there are multiple nonmissing laboratory values for a participant's particular test at a scheduled visit, central laboratory values have higher priority over local laboratory values. If a tie still exists, the laboratory value with the smallest laboratory sequence number will be used in by-visit summaries.

Numeric laboratory values will be summarized descriptively in SI units, and non-numeric test values will be tabulated when necessary. In addition, box-and-whisker plots will be provided for HGB, platelet counts, leukocytes, and neutrophils.

A table for number of participants with abnormal ALT, AST, bilirubin, and creatinine will be tabulated by treatment group.

For test results that will be summarized with available normal ranges, the number and percentage of participants with laboratory values being low (but never high), normal, high (but never low), and both low and high will be calculated for each test. This shift summary will be produced for each test for the safety population. The denominator for the percentage calculation will use the number of participants in the baseline category (ie, low, high, normal, missing) as the denominator for the percentage in each of the categories during the study.

Severity grades will be assigned to laboratory test values based on the numerical component of CTCAE v5. Shift tables will be presented showing change in CTCAE grade from baseline to worst grade postbaseline. Separate summaries for abnormally high and abnormally low laboratory values will be provided when the laboratory parameter has both high and low grading criteria. The denominator for the percentage calculation will be the number of participants in the baseline category. The number of participants who experienced worsening of laboratory abnormalities will be summarized by maximum severity.

In cases where differentials of hematology parameters are obtained without corresponding absolute count data, efforts will be made to investigate if the conversion to an absolute value will lead to additional abnormalities. This will be discussed with the clinical team regarding appropriate documentation and action.

9.3.3. Potential Hy's Law Events

Participants with elevated ALT or AST > 3 × ULN range and alkaline phosphatase < 2 × ULN range accompanied by total bilirubin > 2 × ULN range at the same visit will be listed by treatment group.

9.4. Vital Signs

Values at each scheduled visit, change, and percentage change from baseline for vital signs, including systolic blood pressure, diastolic blood pressure, pulse, temperature, respiratory rate, and weight will be summarized descriptively.

Normal ranges for vital sign values are defined in Table 4. For participants exhibiting vital sign abnormalities, the abnormal values will be listed along with their assigned treatment group. Alert vital signs are defined as an absolute value outside the defined normal range and percentage change greater than 25%. Note that the definition of alert vital signs does not apply for body temperature and weight. The abnormal values for participants exhibiting alert vital sign abnormalities will be listed.

Table 4: Normal Ranges for Vital Sign Values

Parameter	High Threshold	Low Threshold
Systolic blood pressure	≤ 155 mmHg	≥ 85 mmHg
Diastolic blood pressure	≤ 100 mmHg	≥ 40 mmHg
Pulse	≤ 100 bpm	≥ 45 bpm
Temperature	≤ 38°C	≥ 35.5°C
Respiratory rate	≤ 24 breaths/min	≥ 8 breaths/min

9.5. Electrocardiograms

Twelve-lead ECGs including PR, QT, QRS, RR, and QTc intervals will be obtained for each participant during the study. Values at each scheduled visit, change, and percentage change from baseline will be summarized for each ECG parameter. Baseline will be determined as the average of all nonmissing values before the first administration of parsaclisib, placebo or ruxolitinib.

Normal ranges for ECG values are defined in Table 5. ECG values will also be considered abnormal if the absolute percentage change from baseline is more than 25% (30% for QRS interval). Participants exhibiting ECG abnormalities will be listed with study visit and by assigned treatment group. Abnormal values for participants with alert ECG values, defined as both the absolute value and the percentage change from baseline being outside normal ranges, will be identified and listed. Outliers of QT and QTc values, defined as absolute values ≥ 500 milliseconds, ≥ 480 milliseconds, and ≥ 450 milliseconds for each parameter, or change from baseline ≥ 30 milliseconds and ≥ 60 milliseconds for each parameter, will be summarized.

Table 5: Normal Ranges for Electrocardiogram Intervals

Parameter	High Threshold	Low Threshold
PR	≤ 220 ms	≥ 75 ms
QT	≤ 500 ms	≥ 300 ms
QRS	≤ 120 ms	≥ 50 ms
QTc	≤ 450 ms	≥ 295 ms
RR	≤ 1330 ms	≥ 600 ms

Twelve-lead ECGs will be obtained for each participant during the study. Electrocardiogram abnormalities, both at baseline and postbaseline visits, will be tabulated by treatment group. Incidences of abnormalities and a description of the abnormality will be listed with study visit and assigned treatment group.

10. INTERIM ANALYSES

10.1. Overview of Interim Analyses

There will be 2 planned interim analyses conducted for this study. The first interim analysis will be conducted when the first 100 of the planned randomized participants (approximately 50 in each treatment group) reach Week 12 assessments of spleen volume and MFSAF TSS or discontinue from treatment. The first interim analysis is to test Week 12 spleen volume and MFSAF TSS for futility. An HSD (Hwang et al 1990) with $\gamma = 0$ will be used to determine the nonbinding futility boundary for spleen volume and MFSAF TSS. The study will be terminated early for futility if either the spleen volume or the MFSAF TSS endpoints cross the futility boundary. Based on the results of this interim analysis, the DMC may recommend to terminate the study, continue the study with no changes to enrollment, or continue the study with no further enrollment permitted until review of the second interim analysis.

The second interim analysis will be conducted when the first 220 randomized participants (approximately 110 in each treatment group) reach Week 12 assessments of spleen volume and MFSAF TSS or discontinue from treatment. The second interim analysis is for the efficacy check at 1-sided alpha of 10%. The study will continue if both spleen volume and MFSAF TSS endpoints are positive. The DMC may recommend to either terminate the study for lack of efficacy or continue the study, based on the results of this analysis.

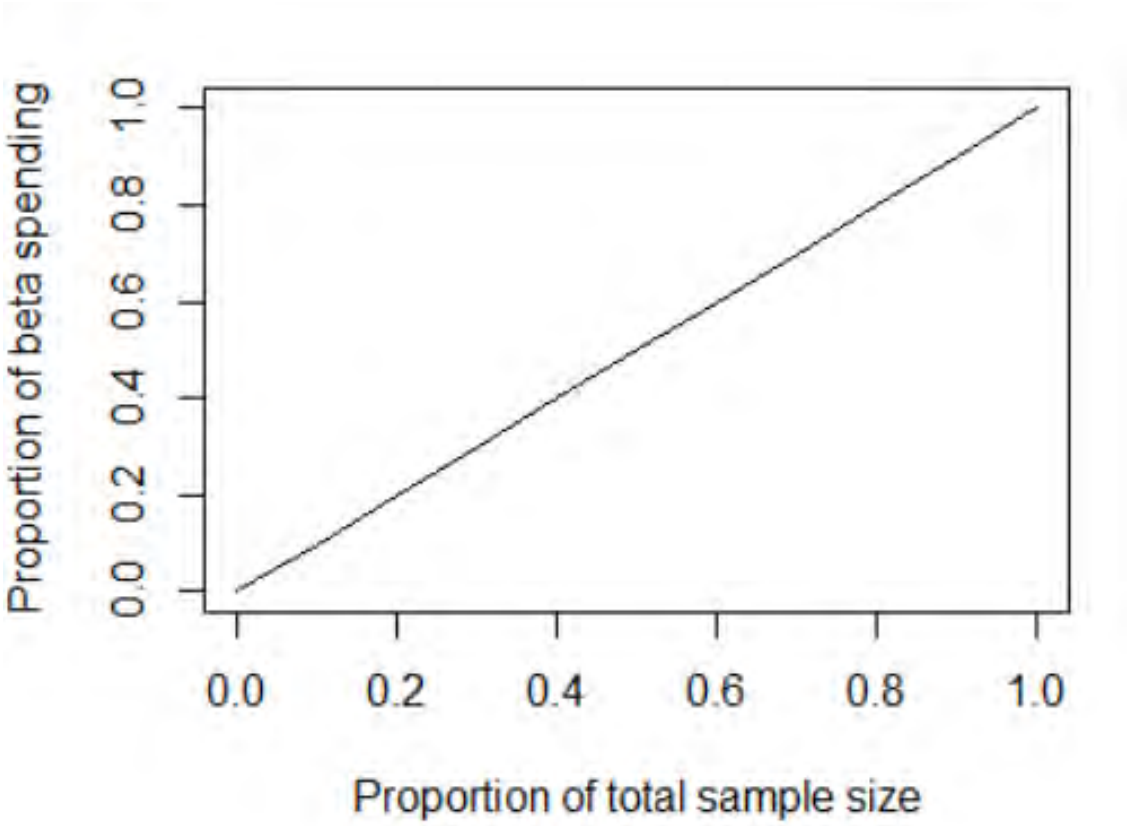
At both interim analyses, MFSAF TSS will be tested only if spleen volume passes the futility/efficacy check.

10.2. Derivations and Calculations for Interim Analyses

For the first analysis of futility, the gamma spending function with spending parameter equal to 0 will be used to allocate beta-spending based on a Type II error of 0.16 for spleen volume reduction and 0.2 for MFSAF TSS under the alternative hypothesis. Figure 1 shows that the beta spending function allocates minimal beta for the first interim analysis.

As no formal decision for efficacy will be made based on the efficacy check in the second interim analysis, no alpha will be spent.

Figure 1: Beta Spending



10.3. Guidelines for Decision Rule

Table 6 and Table 7 provide the projected stopping rules if the interim analyses are conducted at the projected number of participants.

Table 6: Interim Analysis for Spleen Volume With HSD (0)

Number of Participants	Interim Analysis 1		Interim Analysis 2		Final Analysis	
	100		220		440	
Decision Outcome	**Stop for Futility**	**Continue to Test MFSAF TSS**	**Stop for Lack of Efficacy**	**Continue to Test MFSAF TSS**	**Do Not Reject Null Hypothesis**	**Reject Null Hypothesis**
Z-statistic	≤ 0.18	> 0.18	≤ 1.28	> 1.28	≤ 1.96	> 1.96
One-sided p-value	≥ 0.43	< 0.43	≥ 0.10	< 0.10	≥ 0.025	< 0.025
Parsaclisib + ruxolitinib response rate[a] for low platelet	≤ 17.8%	> 17.8%	≤ 23.0%	> 23.0%	≤ 23.4%	> 23.4%
Parsaclisib + ruxolitinib response rate[a] for normal platelet	≤ 42.0%	> 42.0%	≤ 50.0%	> 50.0%	≤ 50.5%	> 50.5%

[a] Assumes spleen volume response rate for 50 to < 100 $\times 10^9$/L inclusive or ≥ 100 $\times 10^9$/L platelet is 17.2% and 41.0% in placebo plus ruxolitinib.

Table 7: Interim Analysis for MFSAF TSS With HSD (0)

Number of Participants	Interim Analysis 1		Interim Analysis 2		Final Analysis	
	100		220		440	
Decision Outcome	**Stop for Futility**	**Continue to Interim Analysis 2**	**Stop for Lack of Efficacy**	**Continue to the Final Analysis**	**Do Not Reject Null Hypothesis**	**Reject Null Hypothesis**
Z-statistic	≤ 0.18	> 0.18	≤ 1.28	> 1.28	≤ 1.96	> 1.96
One-sided p-value	≥ 0.43	< 0.43	≥ 0.10	< 0.10	≥ 0.025	< 0.025
Parsaclisib + ruxolitinib response rate[a] for low platelet	≤ 35.4%	> 35.4%	≤ 42.8%	> 42.8%	≤ 43.6%	> 43.6%
Parsaclisib + ruxolitinib response rate[a] for normal platelet	≤ 46.9%	> 46.9%	≤ 54.6%	> 54.6%	≤ 55.5%	> 55.5%

[a] Assumes TSS response rate for 50 to < 100 $\times 10^9$/L inclusive or ≥ 100 $\times 10^9$/L platelet is 34.5% and 45.9% in placebo plus ruxolitinib.

11. CHANGES AND MODIFICATIONS TO THE ANALYSIS PLAN

All versions of the SAP are listed in Table 8.

Table 8: Statistical Analysis Plan Versions

SAP Version	Date
Original	15 DEC 2020
Amendment 1	26 APR 2022

11.1. Changes to Protocol-Defined Analyses

Not applicable.

11.2. Changes to the Statistical Analysis Plan

11.2.1. Amendment 1

- In Section 4.1.4, the derivation of baseline for participants not treated with any study drug was added.
- In Section 5.3.1, it was clarified that participants with missing baseline assessments will be excluded from the ITT population for the responder analyses.
- In Section 7.1, the combination of DIPSS stratification levels at the time of interim analyses were added.
- In Section 7.3.1 and Section 7.4.2.2, the nonevaluable definitions for the primary endpoint analysis of spleen volume and secondary endpoint analysis of MFSAF TSS, respectively, were added.
- In Section 7.4.2, the window for deriving MFSAF TSS at Week 24 was modified to capture the assessments from the last week before participants return the device.
- [redacted].
- Incorporation of administrative changes. Other minor, administrative changes and changes according to the Protocol Amendment 1 were incorporated throughout the SAP.

12. REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29-41.

Efron B. The efficiency of Cox's likelihood function for censored data. J Am Stat Assoc 1977;72:557-565.

[redacted]

Hwang IK, Shih WJ, De Cani JS. Group sequential designs using a family of Type I error probability spending functions. Stat Med 1990;9:1439-1445.

Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer-Verlag. 1997.

[redacted]

APPENDIX A. PLANNED TABLES, FIGURES, AND LISTINGS

This appendix provides a list of the planned tables, figures, and listings for the Clinical Study Report. Shells are provided in a separate document for tables that are not in the Standard Safety Tables v1.9.

The lists of tables, figures, and listings are to be used as guidelines. Modifications of the lists that do not otherwise affect the nature of the analysis will not warrant an amendment to the SAP.

Tables

Table No.	Title	Population	Standard
Baseline and Demographic Characteristics			
1.1 Disposition			
1.1.1	Analysis Populations	ITT	X
1.1.2	Summary of Participant Disposition	ITT	X
1.1.3	Summary of Number of Participants Enrolled by Country and Site	ITT	X
1.1.4	Summary of Protocol Deviations	ITT	X
1.2 Demography and Baseline Characteristics			
1.2.1	Summary of Demographics and Baseline Characteristics	ITT	X
1.3 Baseline Disease Characteristics			
1.3.1	Summary of Baseline Disease Characteristics	ITT	
1.3.2	Summary of Myelofibrosis History	ITT	
1.3.3	Summary of Myelofibrosis Screening Symptom Score	ITT	
1.3.4	Summary of Prior Systemic Cancer Therapy by WHO Drug Class and Preferred Term	ITT	X
1.4 Prior Medication and Concomitant Medication			
1.4.1	Summary of Prior Medications	ITT	X
1.4.2	Summary of Concomitant Medications	ITT	X
1.5+ Others			
1.5.1	Summary of General Medical History	ITT	X
Efficacy			
2.1 Primary Efficacy			
2.1.1.1	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24	ITT	
2.1.1.2	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24	PP	
2.1.2.1	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Sex	ITT	
2.1.2.2	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Age Group	ITT	
2.1.2.3	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by MF Type	ITT	
2.1.2.4	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Presence/Absence of V617F Mutation at Baseline	ITT	

Table No.	Title	Population	Standard
2.1.2.5	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Presence/Absence of HMR Mutations at Baseline	ITT	
2.1.2.6	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Baseline Spleen Volume Group	ITT	
2.1.2.7	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Baseline Spleen Palpation Size Group	ITT	
2.1.2.8	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Race	ITT	
2.1.2.9	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Ethnicity	ITT	
2.1.2.10	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by Platelet Count	ITT	
2.1.2.11	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 35% Reduction in Spleen Volume at Week 12 and Week 24 – by DIPSS	ITT	
2.2 Secondary Efficacy			
2.2.1.1	Time to the First ≥ 35% Reduction in Spleen Volume	ITT	
2.2.1.2	Duration of Maintenance of a ≥ 35% Reduction in Spleen Volume	ITT With Spleen Volume Response	
2.2.1.3	Time From the First ≥ 35% Reduction in Spleen Volume to Loss of ≥ 35% Reduction in Spleen Volume With a ≥ 25% Spleen Volume Increase From Nadir	ITT With Spleen Volume Response	
2.2.1.4	Summary of Spleen Volume by Visit	ITT	
2.2.2.1	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 50% Reduction in MFSAF v4.0 Total Symptom Score at Week 12 and Week 24	ITT	
2.2.2.2	Cochran-Mantel-Haenszel Test of Proportion of Participants With ≥ 50% Reduction in MFSAF v4.0 Total Symptom Score at Week 12 and Week 24	PP	
2.2.2.3	Time to the First ≥ 50% Reduction in MFSAF v4.0 Total Symptom Score	ITT	
2.2.2.4	Summary of MFSAF v4.0 Total Symptom Score Using 7-Day Window by Visit	ITT	
2.2.2.5.1	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Fatigue	ITT	
2.2.2.5.2	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Night Sweats	ITT	
2.2.2.5.3	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Itchiness	ITT	
2.2.2.5.4	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Abdominal Discomfort	ITT	

Table No.	Title	Population	Standard
2.2.2.5.5	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Pain Under Left Ribs	ITT	
2.2.2.5.6	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Early Satiety	ITT	
2.2.2.5.7	Summary of MFSAF v4.0 Individual Symptom Scores by Visit – Bone Pain	ITT	
2.2.2.6	Summary of MFSAF v4.0 Spleen-Related Symptom Scores by Visit	ITT	
2.2.2.7.1	Overall Survival	ITT	
2.2.2.7.2	Overall Survival	PP	
2.2.2.7.3	Overall Survival Adjusted by Rank Preserving Structural Failure Time	ITT	
[redacted]			

Table No.	Title	Population	Standard
[redacted]			
Safety			
3.1 Dose Exposure			
3.1.1	Summary of Exposure and Duration of Exposure to Parsaclisib/Placebo	Safety	X
3.1.2	Summary of Exposure and Duration of Exposure to Ruxolitinib	Safety	X
3.1.3.1	Summary of Parsaclisib Compliance	Safety	X
3.1.3.2	Summary of Ruxolitinib Compliance	Safety	X
3.2 Adverse Events			
3.2.1	Overall Summary of Treatment-Emergent Adverse Events	Safety	X
3.2.2	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.3	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency	Safety	X
3.2.4	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class, Preferred Term, and CTCAE Grade Category	Safety	X
3.2.5	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.6	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency	Safety	X
3.2.7	Summary of Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.8	Summary of Serious Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency	Safety	X
3.2.9.1	Summary of Parsaclisib Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.9.2	Summary of Ruxolitinib Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.10.1	Summary of Parsaclisib Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency	Safety	X
3.2.10.1	Summary of Ruxolitinib Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency	Safety	X
3.2.11.1	Summary of Parsaclisib Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class, Preferred Term, and CTCAE Grade Category	Safety	X
3.2.11.2	Summary of Ruxolitinib Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class, Preferred Term, and CTCAE Grade Category	Safety	X

Table No.	Title	Population	Standard
3.2.12.1	Summary of Grade 3 or Higher Parsaclisib Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.12.2	Summary of Grade 3 or Higher Ruxolitinib Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.13.1	Summary of Parsaclisib Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.13.2	Summary of Ruxolitinib Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.14	Summary of Treatment-Emergent Adverse Events With a Fatal Outcome by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.15.1	Summary of Treatment-Emergent Adverse Events Leading to Parsaclisib Dose Reduction by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.15.2	Summary of Treatment-Emergent Adverse Events Leading to Ruxolitinib Dose Reduction by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.16.1	Summary of Treatment-Emergent Adverse Events Leading to Parsaclisib Dose Interruption by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.16.2	Summary of Treatment-Emergent Adverse Events Leading to Ruxolitinib Dose Interruption by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.17.1	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Parsaclisib by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.17.2	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Ruxolitinib by MedDRA System Organ Class and Preferred Term	Safety	X
3.2.18	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term: Life-Table Method	Safety	X
3.2.19	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term: Life-Table Method	Safety	X
3.2.20.1	Summary of Treatment-Emergent Adverse Events of Special Interest by MedDRA Preferred Term	Safety	X
3.2.20.2	Kaplan-Meier Estimates of Time to the First Onset of Treatment-Emergent Adverse Events of Special Interest by MedDRA Preferred Term	Safety	X
3.2.20.3	Kaplan-Meier Estimates of Longest Duration of Treatment-Emergent Adverse Events of Special Interest by MedDRA Preferred Term	Safety	X
3.3 Laboratory			
3.3.1.1	Summary of Laboratory Values - Hematology	Safety	X
3.3.1.2	Summary of Laboratory Values - Chemistry	Safety	X
3.3.1.3	Summary of Laboratory Values - Coagulation	Safety	X
3.3.1.4	Summary of Laboratory Values - Urinalysis	Safety	X
3.3.2.1	Shift Summary of Hematology Values - to the Worst Abnormal Value	Safety	X
3.3.2.2	Shift Summary of Chemistry Values - to the Worst Abnormal Value	Safety	X

Table No.	Title	Population	Standard
3.3.2.3	Shift Summary of Coagulation Values - to the Worst Abnormal Value	Safety	X
3.3.3.1	Shift Summary of Hematology Laboratory Values in CTC Grade - to the Worst Abnormal Value	Safety	X
3.3.3.2	Shift Summary of Chemistry Laboratory Values in CTC Grade - to the Worst Abnormal Value	Safety	X
3.3.3.3	Shift Summary of Coagulation Laboratory Values in CTC Grade - to the Worst Abnormal Value	Safety	X
3.3.3.4	Treatment-Emergent Worsening of Laboratory Abnormalities - Hematology	Safety	X
3.3.3.5	Treatment-Emergent Worsening of Laboratory Abnormalities - Chemistry	Safety	X
3.3.3.6	Treatment-Emergent Worsening of Laboratory Abnormalities - Coagulation	Safety	X
3.3.4.1	Summary of Participants With Elevations in Liver Chemistry Tests	Safety	X
3.4 Vital Signs			
3.4.1	Summary of Systolic Blood Pressure	Safety	X
3.4.2	Summary of Diastolic Blood Pressure	Safety	X
3.4.3	Summary of Pulse	Safety	X
3.4.4	Summary of Respiratory Rate	Safety	X
3.4.5	Summary of Body Temperature	Safety	X
3.4.6	Summary of Weight	Safety	X
3.5 ECG			
3.5.1	Summary of PR Interval (ms) From 12-Lead ECG	Safety	X
3.5.2	Summary of QRS Interval (ms) From 12-Lead ECG	Safety	X
3.5.3	Summary of QT Interval (ms) From 12-Lead ECG	Safety	X
3.5.4	Summary of QTc Interval (ms) From 12-Lead ECG	Safety	X
3.5.5	Summary of Outliers of QT and QTc Interval Values (ms) From 12-Lead ECG	Safety	X
3.5.6	Summary of Clinically Significant ECG Abnormality	Safety	X

Figures

Figure No.	Title
4.2 Secondary Efficacy	
4.2.1.1	Kaplan-Meier Estimates of Time to the First ≥ 35% Reduction in Spleen Volume
4.2.1.2	Kaplan-Meier Estimates of Duration of Maintenance of a ≥ 35% Reduction in Spleen Volume
4.2.2	Kaplan-Meier Estimates of Time to the First ≥ 50% Reduction in MFSAF v4.0 Total Symptom Score
4.2.3	Kaplan-Meier Estimates of Overall Survival
4.2.4	Plot of Proportion of Participants With ≥ 25% Reduction in Spleen Volume at Each Visit
4.2.5	Plot of Percentage Change From Baseline in MFSAF v4.0 Total Symptom Score at Each Visit
4.2.6	Plot of Percentage Change From Baseline in Spleen Length at Each Visit
[redacted]	
4.6 Laboratory Data	
4.6.1	Box-and-Whisker Plot of Selected Laboratory Values by Study Visit

Listings

Listing No.	Title
2.1 Discontinued Participants (Participant Disposition)	
2.1.1	Participant Enrollment and Disposition Status
2.1.2	Participant Inclusion and Exclusion Criteria Violations
2.2 Protocol Deviations	
2.2.1	Protocol Deviations
2.3 Data Excluded From PK, Efficacy, and/or Safety Analyses	
2.3.1	Analysis Population
2.3.2	Crossover Participants
2.4 Demographic and Baseline Characteristics (Including Prior and Concomitant Medications)	
2.4.1	Demographic and Baseline Characteristics
2.4.2	Disease History
2.4.3	Prior Radiation Treatment
2.4.4	Prior Systemic Therapy for MF Other Than Ruxolitinib
2.4.5	Prior Surgery or Surgical Procedure
2.4.6	Medical History
2.4.7	Prior and Concomitant Medication
2.5 Drug Compliance	
2.5.1	Study Drug Compliance of Parsaclisib
2.5.2	Study Drug Compliance of Ruxolitinib
2.5.3	Study Drug Administration of Parsaclisib/Placebo
2.5.	Study Drug Administration of Ruxolitinib
2.6 Efficacy (and/or PK Data)	
2.6.1	Spleen Volume
2.6.2	MFSAF v4.0 Total Symptom Score
2.6.3	Death
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
2.7 Adverse Events	
2.7.1	Adverse Events
2.7.2	Dose-Limiting Toxicities
2.7.3	Serious Adverse Events
2.7.4	Grade 3 and Higher Adverse Events
2.7.5	Fatal Adverse Events
2.7.6	Treatment-Related Adverse Events

Listing No.	Title
2.7.7	Adverse Events Leading to Interruption, Reduction, or Discontinuation of Parsaclisib or Ruxolitinib
2.8 Laboratory Data	
2.8.1	Clinical Laboratory Values – Hematology
2.8.2	Clinical Laboratory Values – Chemistry
2.8.3	Abnormal Clinical Laboratory Values
2.8.4	Potential Hy's Law Events
2.8.5	Central Laboratory Collection Times
2.9 Vital Signs	
2.9.1	Vital Signs
2.9.2	Abnormal Vital Sign Values
2.9.3	Alert Vital Sign Values
2.10 ECG	
2.10.1	12-Lead ECG Values
2.10.2	Abnormal 12-Lead ECG Values
2.10.3	Alert 12-Lead ECG Values
2.11 Physical Examination	
2.11.1	Physical Examinations

[redacted]