NOVARTIS

Clinical Development

INC424/Ruxolitinib

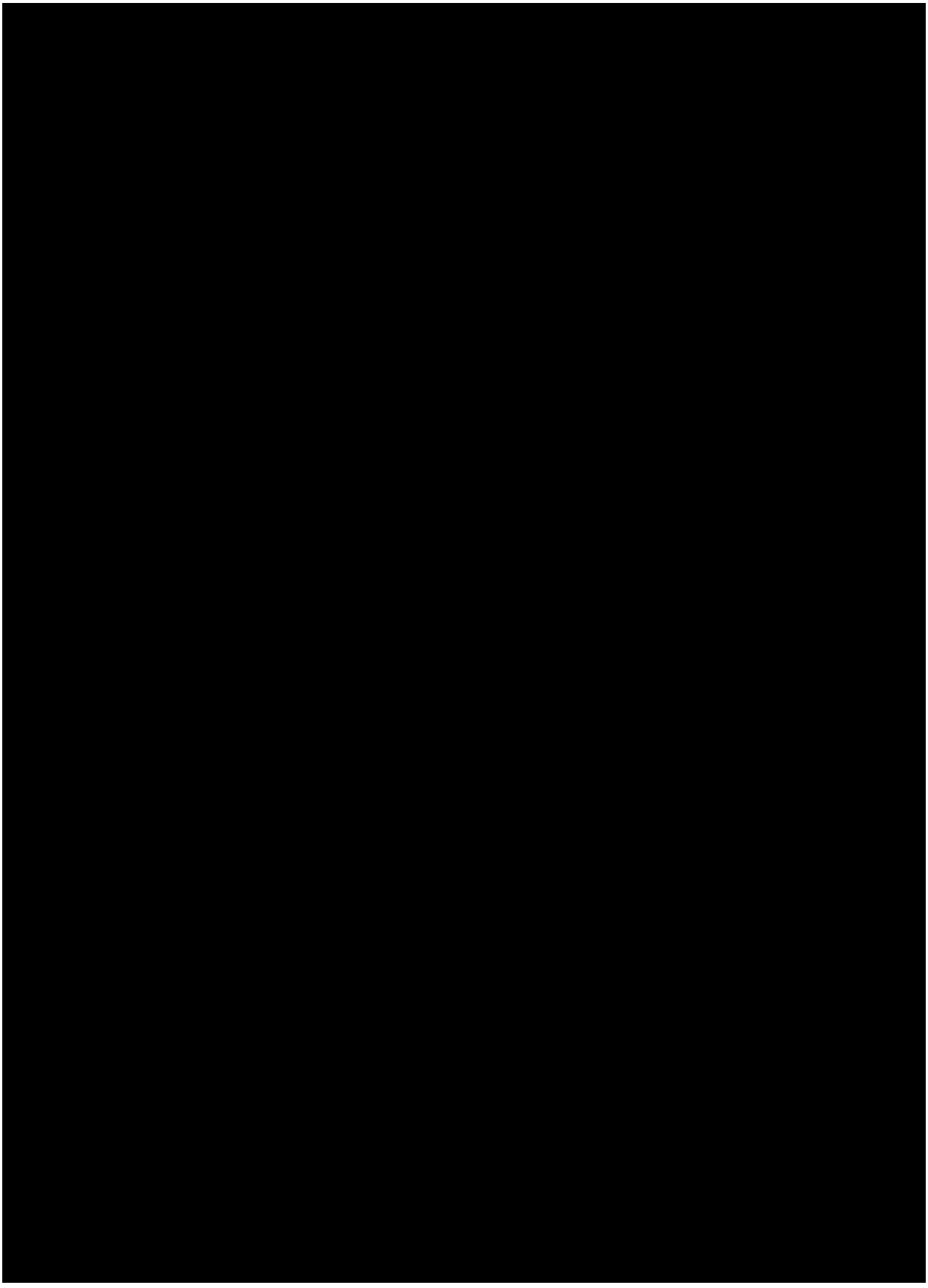
CINC424J12301 / NCT04362137

Phase 3 randomized, double-blind, placebo-controlled, multi-center study to assess the efficacy and safety of ruxolitinib in patients with COVID-19 associated cytokine storm (RUXCOVID)

Statistical Analysis Plan (SAP)

Author:	Trial Statistician, [redacted]

Document type:	SAP Documentation
Document status:	Amendment 1
Release date:	24-Nov-2020
Number of pages:	32



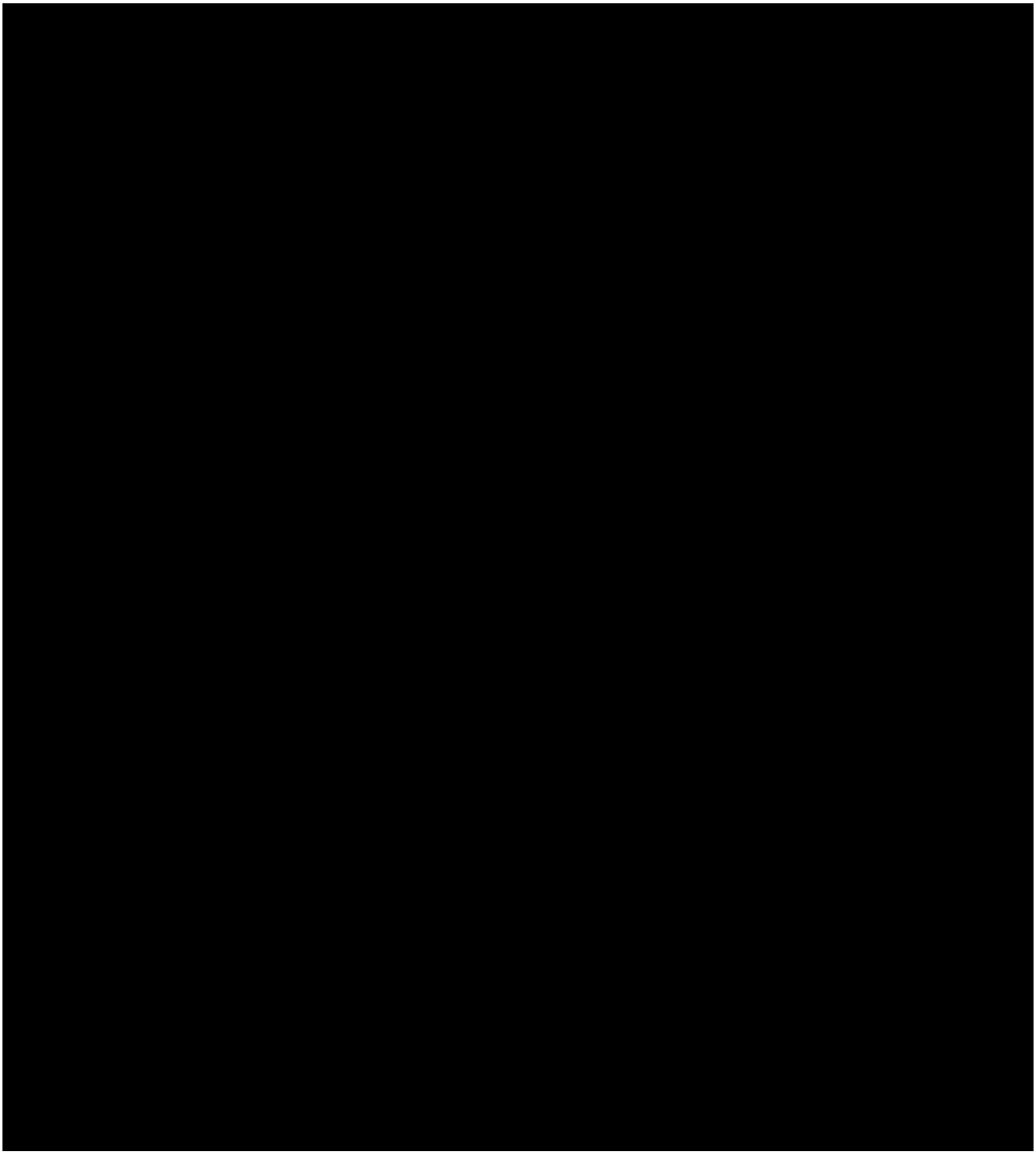


Table of contents

List of tables

List of figures

List of abbreviations

AE	Adverse event
AESI	Adverse event of special interest
ANCOVA	Analysis of covariance
ATC	Anatomical Therapeutic Classification
BMI	Body mass index
[redacted]	[redacted]
CRS	Case Retrieval Strategy
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
ICU	Intensive Care Unit
IRT	Interactive Response Technology
J2R	Jump-to-reference
MAR	Missing at random
MedDRA	Medical Dictionary for Drug Regulatory Affairs
NEWS2	National Early Warning Score
PCR	Polymerase chain reaction
PD	Protocol deviation
PDS	Programming Dataset Specification
POM	Proportional odds model
PT	Preferred Term
RAS	Randomized Set
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SoC	Standard-of-Care
SOC	System Organ Class
WHO	World Health Organization

1 Introduction

This document contains details of the statistical methods that will be used in the phase III clinical trial CINC424J12301. This study is designed to evaluate the efficacy and safety of ruxolitinib in the treatment of patients with COVID-19 pneumonia. The primary endpoint is a composite endpoint (also referred as "clinical failure" in the protocol) defined as the proportion of patients who die, develop respiratory failure (require mechanical ventilation), or require intensive care unit (ICU) care by Day 29.

Data will be analyzed according to Section 12 of the study protocol.

The following document was referred while writing the SAP:

CINC424J12301 Clinical Trial Protocol Final version 01 dated 20-May-2020.

Important information is given in the following sections and details are provided, as applicable, in Section 5: Appendix.

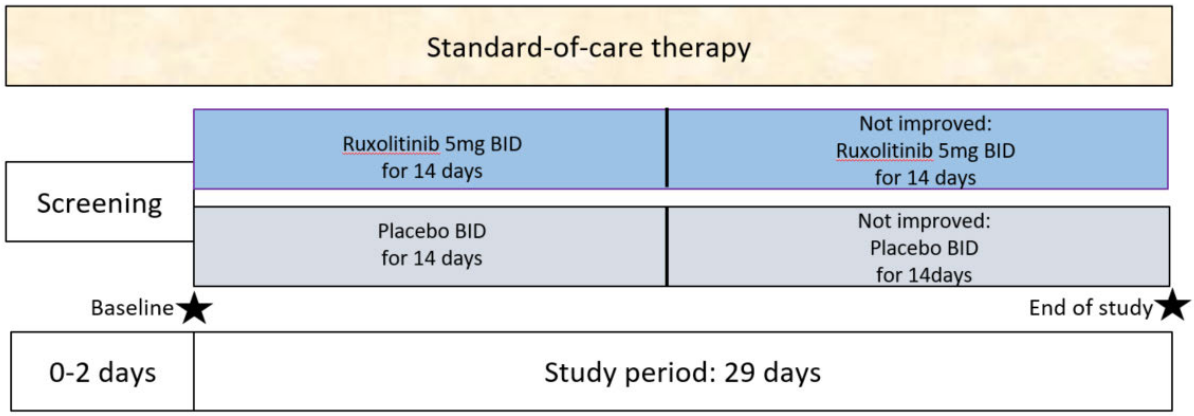
1.1 Study design

This is a randomized, double-blind, placebo-controlled, 29-day, multicenter study to assess the efficacy and safety of ruxolitinib + SoC therapy, compared with placebo + SoC therapy, in patients aged ≥12 years with COVID-19-induced pneumonia (Figure 1-1).

The study will include:

- **Screening period** of 0-2 days
- **Study period** of 29 days (treatment of 14 days; an additional 14 days of study drug may be given, if in the opinion of the investigator, the patient's clinical signs and symptoms are neither improved nor worsened and the potential benefit outweighs the potential risk) – see below.

Figure 1-1 Study design



Eligible patients will be randomized on the same day as screening or up to two days after completing the screening procedures. At Day 1 (randomization visit), patients will be assigned in a 2:1 ratio to receive oral ruxolitinib 5 mg twice daily or oral matching-image placebo for a total of 14 days. An additional 14 days of study drug may be given, if in the opinion of the

investigator, the patient's clinical signs and symptoms are neither improved nor worsened and the potential benefit outweighs the potential risk.

Approximately 402 patients in total are needed (268 randomized to ruxolitinib and 134 to placebo). Detailed information regarding sample size calculation is provided in Section 3.

Randomization will be stratified by geographic region (North America, West Europe, East Europe, Latin America, and Other).

The primary endpoint will be assessed over 29 days.

No interim analysis is planned. A Data Monitoring Committee (DMC) at Novartis will be established to conduct periodical unblinded safety reviews. The analysis plan for the DMC will be created separately.

1.2 Study objectives and endpoints

Section 2 of the study protocol lists the following primary, secondary, [redacted] objectives.

Table 1-1 Objectives and related endpoints

	Objectives	Endpoints
Primary Objective	To evaluate the efficacy (as measured by a composite endpoint of proportion of patients who die, develop respiratory failure (require mechanical ventilation), or require intensive care unit (ICU) care) of ruxolitinib + standard-of-care (SoC) therapy compared with placebo + SoC therapy, for the treatment of COVID-19 by Day 29.	Composite endpoint defined as: • Death **OR** • Respiratory failure (require mechanical ventilation) **OR** • Intensive care unit (ICU) care by Day 29.
Secondary Objectives	To evaluate the efficacy (as measured by clinical status using a 9-point ordinal scale) of ruxolitinib + SoC therapy compared with placebo + SoC therapy, for the treatment of COVID-19 (WHO 18-Feb-2020).	Clinical status assessed using a 9-point ordinal scale (Section 2.7.1.1) at Day 15 and Day 29. • Percentage of patients with a better category (lower number) in clinical status at Day 15 and at Day 29. • Percentage of patients with at least two-point improvement in clinical status at Day 15 and at Day 29. • Percentage of patients with at least one-point improvement in clinical status at Day 15 and at Day 29. • Percentage of patients with at least one-point deterioration in clinical status at Day 15 and at Day 29. • Time to improvement from baseline category to one less severe category of the ordinal scale. • Mean change in the 9-point ordinal scale from baseline to Days 15 and 29.

	Objectives	Endpoints
	To evaluate the efficacy of ruxolitinib + SoC therapy, compared with placebo + SoC therapy, on in-hospital outcomes in patients with COVID-19.	Mortality rate at Day 15 and at Day 29; Proportion of patients requiring mechanical ventilation by Day 29; Duration of hospitalization.
	To evaluate the efficacy of ruxolitinib + SoC therapy, compared with placebo + SoC therapy, in change in the National Early Warning Score (NEWS2) score in patients with COVID-19.	The time to discharge or to a NEWS2 (Appendix 3 of the protocol) score of ≤2 and maintained for 24 hours whichever comes first. Change from baseline to Days 3, 5, 8, 11, 15, and 29 in NEWS2 score.
	To evaluate the efficacy of ruxolitinib + SoC therapy, compared with placebo + SoC therapy, in change in SpO_2/FiO_2 ratio in patients with COVID-19.	Change from baseline to Days 15 and 29 in SpO_2/FiO_2 ratio.
	To evaluate the efficacy of ruxolitinib + SoC therapy, compared with placebo + SoC therapy, in proportion of patients with no oxygen therapy (defined as oxygen saturation ≥ 94% on room air) in patients with COVID-19.	Proportion of patients with no oxygen therapy at Days 15 and 29.
	To evaluate the safety of ruxolitinib + SoC therapy, compared with placebo + SoC therapy, in the treatment of patients with COVID-19.	Number of participants with treatment-related adverse events (as assessed by Common Terminology Criteria for Adverse Event version 5.0), serious adverse events, clinically significant changes in laboratory measures and vital signs during the 29-day Study Period.
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]

Objectives	Endpoints
[redacted]	[redacted]

2 Statistical methods

2.1 Data analysis general information

The statistical analysis will be performed by Novartis. The most recent version of SAS® (SAS Institute Inc., Cary, NC, USA) available in the statistical programming environment of Novartis will be used for the analysis. R version 3.6.1 may also be used as appropriate.

2.1.1 General definitions

The terms 'double-blind treatment' will be used in this document and refer to the double-blind INC424 and placebo.

2.1.1.1 Study day

Study day will be defined as the number of days since the date of randomization. The date of randomization will be defined as Day 1 and the day prior to randomization will be defined as Day -1.

Therefore, for a particular date, study day will be calculated as follows:

for dates on or after the date of randomization,

Study day = Assessment date – Day 1 + 1;

for dates prior to the date of randomization,

Study day = Assessment date – Day 1.

If a patient was never randomized but received at least one dose of double-blind treatment, the date of first dose of double-blind treatment will be used instead of the randomization date. In this case, the date of first dose of double-blind treatment is defined as Day 1 and the day before the first dose of double-blind treatment is defined as Day -1.

2.1.1.2 Baseline definition

In general, baseline is defined as the last measurement before the first dose of double-blind treatment at Day 1. Details on calculation of baseline will be provided in the latter sections.

2.1.1.3 Post-baseline measurement

Post-baseline measurements are defined as those assessments after the first dose of double-blind treatment.

When change from baseline is of interest the following formula will be used for each scheduled visit and time-point where baseline and post-baseline values are both available:

Change from baseline = post-baseline value – baseline value.

If not stated otherwise for efficacy analyses, on-treatment values are defined as values taken post-baseline but no later than 1 day after last dose of double-blind treatment. Off-treatment values are defined as post-baseline values taken more than 1 day after last dose of double-blind treatment. All on- and off- treatment values will be used in the efficacy analyses, unless otherwise specified.

For safety analyses other than adverse events (AEs), e.g. laboratory, vital signs, treatment-emergent values are defined as values taken post-baseline and up to the last study visit (Day 29).

AEs will be considered as treatment-emergent if the event starts after the first dose of double-blind treatment or the event is present prior to start of double-blind treatment but increased in severity based on preferred term and up to the last study visit (Day 29).

Details on calculation of post-baseline values are provided in the latter sections.

2.2 Analysis sets

The Randomized Analysis Set (RAS) consists of all randomized patients. Following the intent-to-treat principle, patients will be analyzed according to the treatment they were assigned to at randomization.

The Safety Set will include all patients who received at least one dose of double-blind treatment. Patients will be analyzed according to the treatment they received, where treatment received is defined as the randomized/assigned treatment if the participant took at least one dose of that treatment or the first treatment received if the randomized/assigned treatment was never received.

In cases where an incorrect randomization stratum is entered in the interactive response technology (IRT) system the corrected stratum information as per the case report form data will be used for reporting and analysis.

2.2.1 Subgroup of interest

Subgroup analyses are conducted to assess consistency of the treatment effect among the subgroups, without multiplicity adjustments. The following subgroups will be evaluated for the primary endpoint, i.e., defined as the proportion of patients who die, develop respiratory failure (require mechanical ventilation), or require ICU care by Day 29:

- Age category (< 18, 18 - < 65, ≥ 65 years)
- Gender (male, female)
- Race (White, Asian, Black, other)
- Ethnicity (Hispanic, non-Hispanic)
- Region (North America, East Europe, West Europe, and Latin America)
- Body mass index (BMI) (≤ 30.0 kg/m^2, > 30.0 kg/m^2)
- Baseline steroid use (Yes, No)

- Baseline clinical status based on 9-point scale (3, 4, 5)
- Baseline hypertension (Yes, No)
- Baseline diabetes (Yes, No)
- [redacted]
- Time between onset of COVID-19 symptoms and randomization (≤ 10, > 10 days)

Baseline steroid use is defined as usage between onset of symptoms and randomization (inclusive). Baseline hypertension/diabetes is defined as "Yes" if patients have the comorbidity ongoing at randomization.

In addition to the primary endpoint, subgroups of patients with different baseline clinical status based on 9-point scale (3, 4, 5) will be evaluated for the following endpoints:

- Mortality by Day 29
- [redacted]

2.3 Patient disposition, demographics and other baseline characteristics

No inferential testing on the differences in patient disposition, demographics and other baseline characteristics between treatment arms will be performed.

2.3.1 Patient disposition

The RAS will be used for the summary and listing of patient disposition. The screening disposition and the analysis sets table will be based on all screened patients.

The number of patients in the RAS will be summarized by region, country, center and treatment group. Further, the overall number of patients who entered, completed, and discontinued study will be summarized including the reasons for discontinuation for each period: pre-randomization, and double-blind treatment. Patients who permanently discontinued from the planned treatment phase will be listed including reason and date of discontinuation.

Number of patients with protocol deviations will be tabulated by category (e.g., selection criteria not met, subject not withdrawn as per protocol, treatment deviation, prohibited concomitant medication, other) and deviation.

The number of patients included in each analysis set will be tabulated. Patients exclusion from analysis sets will be listed for all patients with reasons for exclusion (i.e. including both protocol and non-protocol deviations).

2.3.2 Patient demographics and other baseline characteristics

Demographic and other baseline data including disease characteristics will be listed and summarized descriptively by treatment group for the RAS.

Demographic characteristics, including age, gender, race, ethnicity, country, height, weight, and BMI will be summarized by treatment and overall.

Baseline disease characteristics, including number of days between onset of symptoms and randomization, number of days between diagnosis and randomization, clinical status based on

the 9-category ordinal scale, SpO_2/FiO_2, pneumonia (Yes, No), remdesivir use (Yes, No), and steroid use (Yes, No) will be summarized.

Categorical data will be presented as frequencies and percentages including a categroy for missing data if any. For continuous data, mean, standard deviation, median, minimum, and maximum will be presented. For selected parameters, 25th and 75th percentiles will also be presented.

No statistical analyses will be provided for baseline comparability between the treatment groups.

In addition, the following categorizations of continuous variables will be done:

- Age into < 18, 18 - < 65 years, and ≥ 65 years;
- BMI into ≤ 30.0 kg/m^2 and > 30.0 kg/m^2.

2.3.3 Medical history/current medical condition

Medical history and current medical conditions will be coded with the Medical Dictionary for Regulatory Activities terminology (MedDRA) using the most recent version at the time of database lock. History/conditions, including pre-specified protocol solicited events, will be summarized for the RAS by primary system organ class and preferred term. In addition, the pre-specified protocol solicited events will be summarized be pre-specified medical history term.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

All summaries of treatments will be performed on the Safety Set.

2.4.1 Study treatment / compliance

The duration of exposure in days to ruxolitinib and standard of care therapies will be summarized by means of descriptive statistics. The categorized duration of exposure (≤ 7 days, 8 - ≤ 15 days, 16 - ≤ 21 days, 22 - ≤ 28 days, ≥ 29 days) will also be summarized by treatment group.

Duration of exposure to double-blind treatment will be calculated as the number of days between the first dose date and the last dose date exposed to that treatment over the specified period (expressed as: Duration of exposure = Date of last known dose of double-blind treatment – Date of first dose of double-blind treatment + 1).

The number of patients who permanently discontinued from double-blind treatment and the reasons will be summarized by treatment group.

The number of patients who had any dose modifications/interruptions and the reasons will be summarized by treatment.

2.4.2 Prior, concomitant and post therapies

Summaries will be performed separately for medications prior to screening (medications starting and ending prior to screening), for concomitant medications (medications which were taken anytime between the first dose and last dose of randomized treatment, inclusive), and for

concomitant medications related to study indication. These medications will be listed and summarized according to the Anatomical Therapeutic Chemical (ATC) classification system and PT by treatment group. More than one ATC class per medication is possible and the medication will be reported under all applicable classes. Patients taking any of the tocilizumab, canakinumab, sarilumab or anakinra will also be summarized.

Patients taking prohibited concomitant medications will be noted in the summary of protocol deviations.

Surgical and medical procedures (non-drug therapies) will be listed and summarized by primary system organ class (SOC) and PT by treatment group, separately for prior and concomitant procedures.

2.5 Analysis of the primary objective

The primary objective of the study is to evaluate the efficacy, as measured by a composite endpoint of proportion of patients who die, develop respiratory failure (require mechanical ventilation), or require intensive care unit (ICU) care, of ruxolitinib + standard-of-care (SoC) therapy compared with placebo + SoC therapy, for the treatment of COVID-19 by Day 29. The primary analysis for this study will be conducted using the treatment policy strategy.

2.5.1 Primary endpoint

The primary clinical question of interest is: Does ruxolitinib + SoC decrease the probability of meeting the primary composite endpoint (death, respiratory failure (require mechanical ventilation), or ICU care by Day 29) compared with placebo + SoC in patients with COVID-19-induced pneumonia, regardless of other subsequent clinical interventions?

The justification for this primary estimand is that it captures the clinical outcome of most interest after the assignment of double-blind treamtent, which reflects also any effects of additional subsequent interventions potentially due to such clinical decision. In the ongoing COVID-19 pandemic with evolving treatment guidelines and healthcare system burdens, this primary estimand is deemed better to reflect actual clinical practices.

The estimand framework for the primary objective is defined as below.

- Treatment - ruxolitinib or placebo added to SoC therapy
- Population – Hospitalized patients with coronavirus (SARS-CoV)-2 infection confirmed by polymerase chain reaction (PCR) test or another rapid test from the respiratory tract prior to randomization as described in Section 5.1 of the protocol.
- Endpoint - Primary efficacy endpoint is defined as the proportion of patients who died, developed respiratory failure (require mechanical ventilation – defined as in Table 5-1), or required ICU care over 1-29 days
- Population-level summary - Odds-ratio comparing ruxolitinib added to SoC therapy to placebo added to SoC therapy.
- Intercurrent event - Discontinuation of study treatment; the handling of the event is discussed in Section 2.5.3.

2.5.2 Statistical hypothesis, model, and method of analysis

The statistical hypothesis tested for the primary endpoint is that there is no difference in the proportion of subjects meeting the composite primary endpoint by Day 29 with ruxolitinib + SoC versus placebo + SoC therapy.

Let p_i denote the proportion of patients meeting the primary endpoint for treatment groups j, j = 0, 1 where

- 0 corresponds to placebo + SoC
- 1 corresponds to ruxolitinib + SoC

The following statistical hypothesis will be tested to address the primary objective:

H_0: $p_0 = p_1$, H_1: $p_0 \neq p_1$.

The odds of meeting the primary endpoint will be analyzed by a logistic regression model with treatment group, region, baseline clinical status based on the 9-point ordinal scale (≤ 3, ≥ 4), age, and gender as covariates. The estimated odds ratio, p-values, and 95% confidence intervals will be presented.

The study will be considered positive, if ruxolitinib demonstrates a statistically significant greater reduction in the proportion of patients who die, develop respiratory failure (require mechanical ventilation), or ICU care by Day 29. This implies observing an odds ratio of < 1.

In case of separability (often occurs when the event is rare), Firth's penalized maximum likelihood estimation will be performed to reduce bias in the parameter estimates (Heinze and Schemper 2002; Firth 1993). The estimated odds ratio, p-values and 95% CI (all computed by penalized profile likelihood) will be presented.

The numbers and percentages of patients who met the primary composite endpoint will be summarized by treatment groups.

Mis-randomized patients who had developed respiratory failure and/or required ICU care at randomization (thus meeting the exclusion criteria) are excluded from the analysis.

2.5.3 Handling of missing values/censoring/discontinuations

If there is any missing data over 1-29 days, the patient will be considered as meeting the primary composite endpoint, unless they are in one of the scenarios below:

- There was no occurrence of death, respiratory failure (require mechanical ventilation), nor ICU care in all the available data and patients were discharged from the hospital.
- The last available observed data is from Day 15 or later, and there was no occurrence of death, respiratory failure (require mechanical ventilation), nor ICU care in all the available data.

2.5.4 Supportive analyses

Sensitivity analysis

In order to determine the robustness of the logistic regression model used for the primary analysis, a non-parametric regression model (Koch et al 1998) will also be evaluated using the

same covariates as the logistic regression model in the primary analysis. This methodology has minimal assumptions and the estimated odds ratio, p-values, and 95% confidence intervals will be presented.

A tipping point analysis will be implemented to test the robustness of the primary analysis results to different missing data mechanism. Specifically, all possible combinations of the missing values in the primary endpoint in the ruxolitinib + SoC group and the placebo + SoC group will be varied (responder/non-responder) to see if there is any tipping point. For each combination, the contingency table of treatment * response will be tested using a one-sided Z-test.

The between group difference in the primary endpoint will also be assessed by risk difference using the marginal standardization method, for which the risk difference will be derived from the predicted risks for every patient as if they had received the ruxolitinib + SoC or the placebo + SoC using a logistic regression model (Ge et al 2011). The model will include the same covariates as in the primary analysis. The estimated treatment group-specific risks, risk difference, p-value, and 95% CI will be presented.

The primary analysis will be repeated, excluding patients who were randomized but received no double-blind treatment.

The primary analysis will be repeated, including mis-randomized patients who had developed respiratory failure and/or required ICU care at randomization.

Details are provided in Section 5.4.1.

Supplementary analysis

To assess the effect of classifying patients' primary endpoint with missing data, a multiple imputation based analysis will be conducted.

This analysis quantifies the treatment effect in all randomized patients with an adherence to treatment like we would see in clinical practice. The binary outcome of the primary endpoint will be imputed for patients whose value could not be derived (for patients with missing data who died, or developed respiratory failure (required mechanical ventilation) or had ICU care at any time during the 29 days, their primary endpoint is derivable and will be considered as meeting the primary composite endpoint).

- Missing data in the ruxolitinib + SoC arm will be multiply imputed based on placebo + SoC arm data using jump-to-reference (J2R) assumption.
- Missing data in the placebo + SoC arm will be multiply imputed based on the missing at random (MAR) assumption.

Results will be presented similarly to those of primary analyses.

Details are provided in Section 5.4.1.

Supportive analyses

The primary analysis will be repeated for Day 15. The proportion of patients who died, developed respiratory failure, or received ICU care by Day 15 will be analyzed with the same logistic regression model as for the primary endpoint.

The primary analysis will be repeated for each and combinations of the three events in the composite endpoint: death, respiratory failure (requires mechanical ventilation), ICU care, respiratory failure (or death), and ICU care (or death). Separate logistic regression models will be fit to compare the odds between treatment groups for each of the endpoints. Missing data will be handled similarly as outlined for the primary analysis. The estimated odds ratios, p-values, and 95% CIs will be presented. The analyses will be performed for up to Day 15 and Day 29 respectively. The between group difference in these endpoints will also be assessed by risk difference using the marginal standardization method (Ge et al 2011).

The subgroup analyses (subgroups defined in Section 2.2.1) will be explored for the primary estimand and the same logistic regression model as described for the primary analysis with the additional term of subgroup factor (if not already included in the model) and the interaction term of subgroup and treatment. In case of analyses on subgroups with extremely imbalanced sample sizes, the subgroup levels can be combined, if appropriate, while fitting the analysis model, or Firth's penalized maximum likelihood estimation will be performed. The point estimate and 95% CI for odds ratio for each subgroup will be presented using forest plots.

2.6 Analysis of the key secondary objective

There is no key secondary objective defined in the protocol.

2.7 Analysis of secondary efficacy objectives

2.7.1 Secondary endpoints

No multiplicity adjustment will be carried out for secondary analyses described below.

All analysis of secondary endpoints will be performed on the RAS.

2.7.1.1 9-category ordinal scale

The clinical status of the patient is assessed using a 9-category ordinal scale (WHO 18-Feb-2020). The scale is as follows:

Patient State	Descriptor	Score
Uninfected	No clinical or virological evidence of infection	0
Ambulatory (defined as not in hospital or in hospital and ready for discharge)	No limitation of activities	1
	Limitation of activities	2
Hospitalized Mild Disease	Hospitalized, no oxygen therapy (defined as $SpO_2 \geq$ 94% on room air)	3
	Oxygen by mask or nasal prongs	4
Hospitalized Severe Disease	Non-invasive ventilation or high-flow oxygen	5
	Intubation and mechanical ventilation	6
	Ventilation + additional organ support – pressors, RRT (renal replacement therapy), ECMO (extracorporeal membrane oxygenation)	7

Dead	Death	8

The baseline value is defined as the last assessment of clinical status prior to first dose of double-blind treatment. If missing, the baseline value will be assumed to be missing at random as assessments are performed prior to any knowledge of treatment allocation and will be imputed using the median value of all patients with non-missing baseline values in the RAS.

Descriptive statistics

Summary statistics (mean, median, standard deviation, q1, q3, min, max, frequencies and percentages in each category) of clinical status at baseline, Day 15, and Day 29 as well as change from baseline at Day 15 and Day 29 will be summarized by treatment arm.

The mean clinical status on a daily basis over 29 days will be plotted by treatment arm with the 95% confidence interval. A stacked bar chart will also be plotted to display the distribution of clinical status on a daily basis over 29 days.

No imputation will be done for missing data.

Inferential statistics for the clinical status

The odds of observing a better category (lower number) of clinical status at Day 15 will be analyzed with a proportional odds model (POM). The odds ratio for treatment group (ruxolitinib + SoC therapy vs. placebo + SoC therapy) estimated from the POM can be interpreted as a summary of the odds ratios obtained from separate binary logistic regressions using all possible cutoff points of the ordinal outcome (e.g. the cutoff of level 5 'Non-invasive ventilation or high-flow oxygen' will combine levels 0, 1, 2, 3, and 4 versus combined levels 5, 6, 7 and 8). The assumption of POM is that the effect of treatment is identical across all possible cutoff points of the ordinal outcome. The proportional odds assumption will be checked by a score test. The model will include treatment group, region, age, gender, and baseline clinical status ($\leq 3, \geq 4$) as covariates. The estimated odds ratios, p-values and 95% confidence intervals will be presented. The analysis will be repeated for the data at Day 29.

Missing data at Day 15/29 will not be imputed and thus will be excluded from the analysis.

Change from baseline in clinical status

Responder analysis

The treatment groups will also be compared in terms of at least a one-point improvement, at least a two-point improvement, and at least a one-point deterioration in clinical status at Days 15 and 29 using respective logistic regression models with the same covariates as for the POM. The estimated odds ratios, p-values, and 95% CIs will be presented.

If a patient has missing data at Day 15/29, the patient will be treated as a non-responder in the respective analysis.

Change from baseline

Mean change from baseline in the clinical status on the 9-point ordinal scale to Days 15 and 29 will be analyzed using an analysis of covariance (ANCOVA) model with factors for treatment

group, age, gender, and region, as well as the baseline clinical status as a continuous linear covariate. The estimated treatment differences, p-values, and 95% CIs will be presented.

Missing data at Day 15/29 will not be imputed and thus will be excluded from the analysis.

Time to improvement from baseline category

Time to improvement from baseline category to one less severe category of the ordinal scale will be analyzed using a competing risk analysis framework where death will be treated as a competing risk. Specifically, a proportional hazards model for the subdistribution (Fine and Gray 1999) of time to improvement will be performed. Same covariates will be used as for the primary analysis. Patients who did not achieve improvement and did not die will be censored at their last clinical status assessment date. The estimated hazard ratios, p-values, and and 95% CIs wil be presented. The cumulative incidence functions of improvement and death over study days will be plotted by treatment.

In addition, the median time to improvement by treatment as well as the 95% CI will be estimated using the Kaplan-Meier approach, while dead patients will be censored at the maximum follow-up time in the study.

2.7.1.2 Oxygen saturation

The proportion of patients with no oxygen therapy (defined as having a clinical status score of 0, 1, 2, or 3, or oxygen saturation ≥ 94% when measured on room air) at Days 15 and 29 will be analyzed using logistic regression models with the same covariates as for the primary analyses, with baseline SpO2/FiO2 added. The estimated odds ratios, p-values, and 95% CIs will be presented.

Missing data at Day 15/29 will not be imputed and thus will be excluded from the analysis.

2.7.1.3 In-hospital outcomes

Duration of hospitalization (time to hospital discharge) will be analyzed using a competing risk analysis framework (death will be treated as a competing risk). Specifically, a proportional hazards model for the subdistribution (Fine and Gray 1999) of time to hospital discharge will be performed. Same covariates will be used as for the primary analysis. Patients who were not discharged and did not die will be censored at their last assessment date. The estimated hazard ratios, p-values, and and 95% CIs wil be presented. The cumulative incidence functions of hospital discharge and death over study days will be plotted by treatment.

In addition, the median time to hospital discharge by treatment as well as the 95% CI will be estimated using the Kaplan-Meier approach, while dead patients will be censored at the maximum follow-up time in the study.

The analyses of mortality rates at Day 15 and at Day 29, and proportion of patients requiring mechanical ventilation by Day 29 are described in Section 2.5.4.

The subgroups of patients with different baseline clinical status will be evaluated for mortality using a similar logistic regression model as described for the primary analysis, replacing the baseline clinical status ($\leq 3, \geq 4$) by baseline clinical status as a categorical variable, and adding

the interaction term of baseline clinical status and treatment. The point estimate and 95% CI for odds ratio for each subgroup will be summarized.

2.7.1.4 National Early Warning Score (NEWS2)

Time to discharge or to a National Early Warning Score (NEWS2) of ≤ 2 and maintained for 24 hours (whichever comes first; maintained for 24 hours is defined as score being ≤ 2 for two days in a row as the score is collected once per day) will be analyzed using a competing risk analysis framework (death will be a competing risk). A proportional hazards model for the subdistribution will be performed. Patients who did not achieve discharge, nor had a NEWS2 of <= 2 maintanied for 24 hours, nor died will be censored at their last assessment date. The estimated hazard ratios, p-values, and and 95% CIs wil be presented. The cumulative incidence functions will be plotted by treatment by competing risks. The model will include the same covariates as in the primary analysis, with baseline NEWS2 score added.

In addition, the median time to hospital discharge or to a NEWS2 of ≤ 2 and maintained for 24 hours by treatment as well as the 95% CI will be estimated using the Kaplan-Meier approach, while dead patients will be censored at the maximum follow-up time in the study.

Summary statistics of change from baseline to Days 3, 5, 8, 11, 15, and 29 in the NEWS2 will be provided by treatment. No missing data will be imputed.

2.7.1.5 SpO_2/FiO_2 ratio

Summary statistics of change from baseline to Days 15 and 29 in SpO_2/FiO_2 ratio will be provided by treatment. In addition, the following categorization will be done:

- Severity of acute respiratory distress syndrome (ARDS): normal: > 314; mild (235 - 314); moderate: 150 - 234; severe: < 150

2.8 Safety analyses

Safety summaries include only data from the treatment-emergent period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries).

All safety analysis will be performed on the Safety Set.

2.8.1 Adverse events (AEs)

Adverse events will be summarized by treatment group.

The number (and percentage) of subjects with treatment-emergent adverse events, defined as events start after the first dose of double-blind medications or events present prior to start of double-blind treatment but increase in severity based on preferred term and up to last study visit, will be summarized in the following ways:

- by treatment and preferred term (PT).
- by treatment, SOC and PT.

- by treatment, SOC, PT and maximum severity (based on CTCAE grades, categorized as Grades 3-4, Grade 5, and All grades). For this summary an AE with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

Separate summaries by SOC and PT will be provided for:

- AEs suspected to be study medication related
- deaths
- serious adverse events (SAEs)
- AEs leading to treatment discontinuation
- AEs leading to dose interruptions and/or adjustments

Further SAEs will be summarized by PT.

Unless otherwise specified, SOCs will be sorted alphabetically and, within each SOC, the PTs will be sorted in descending order of frequency in the ruxolitinib + SoC arm. A subject with multiple adverse events within a SOC or PT is only counted once towards the total of the SOC or PT.

Listings will be provided for all AEs, SAEs, AEs leading to treatment discontinuation, and deaths.

2.8.1.1 AE reporting for CT.gov and EudraCT

For the legal requirements of clinicaltrials.gov and EudraCT, two required tables on treatment emergent adverse events which are not serious adverse events with an incidence greater than 2% and on treatment emergent serious adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term on the safety set population.

If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE
- more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.8.1.2 Adverse events of special interest / grouping of AEs

The number and percentage of patients who reported treatment-emergent adverse events of special interest (AESI) will be summarized by risk name, PT, maximum severity (based on CTCAE grades) and treatment group.

In addition, the number and percentage of patients who reported treatment-emergent co-morbid AE known to occur with COVID-19 will be summarized by risk name, PT, maximum severity, and treatment group.

Risk names will be sorted alphabetically and, within each risk name, the PTs will be sorted in descending order of frequency in the ruxolitinib + SoC arm. If a patient reported more than one adverse event with the same PT, the AE will be counted only once. If a patient reported more than one AE within the same risk, the patient will be counted only once at that risk.

The Compound Case Retrieval Strategy (CRS) will be used to determine the MedDRA search criteria to be used to identify AESI and treatment-emergent co-morbid AE known to occur with COVID-19. The most recent list of AESI at the time of database lock will be used.

A listing for all AESI will be provided.

2.8.2 Deaths

Fatal AEs will be summarized and listed as specified in Section 2.8.1.

2.8.3 Laboratory data

Summaries of laboratory data will include treatment-emergent measurements, which are defined as measurements taken post-baseline and up to the last study visit.

[redacted]

Grading of laboratory values will be assigned programmatically as per the Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. The calculation of CTCAE grades will be based on the observed laboratory values only; clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher.

The following summaries will be produced for hematology and chemistry laboratory data (by laboratory parameter and treatment):

- Shift tables using CTCAE grades to compare baseline to the worst treamtent-emergent value.

The baseline value is the last value prior to first dose of double-blind treatment.

All laboratory data will be listed, with CTCAE grades by treatment group, patient, and visit/time.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

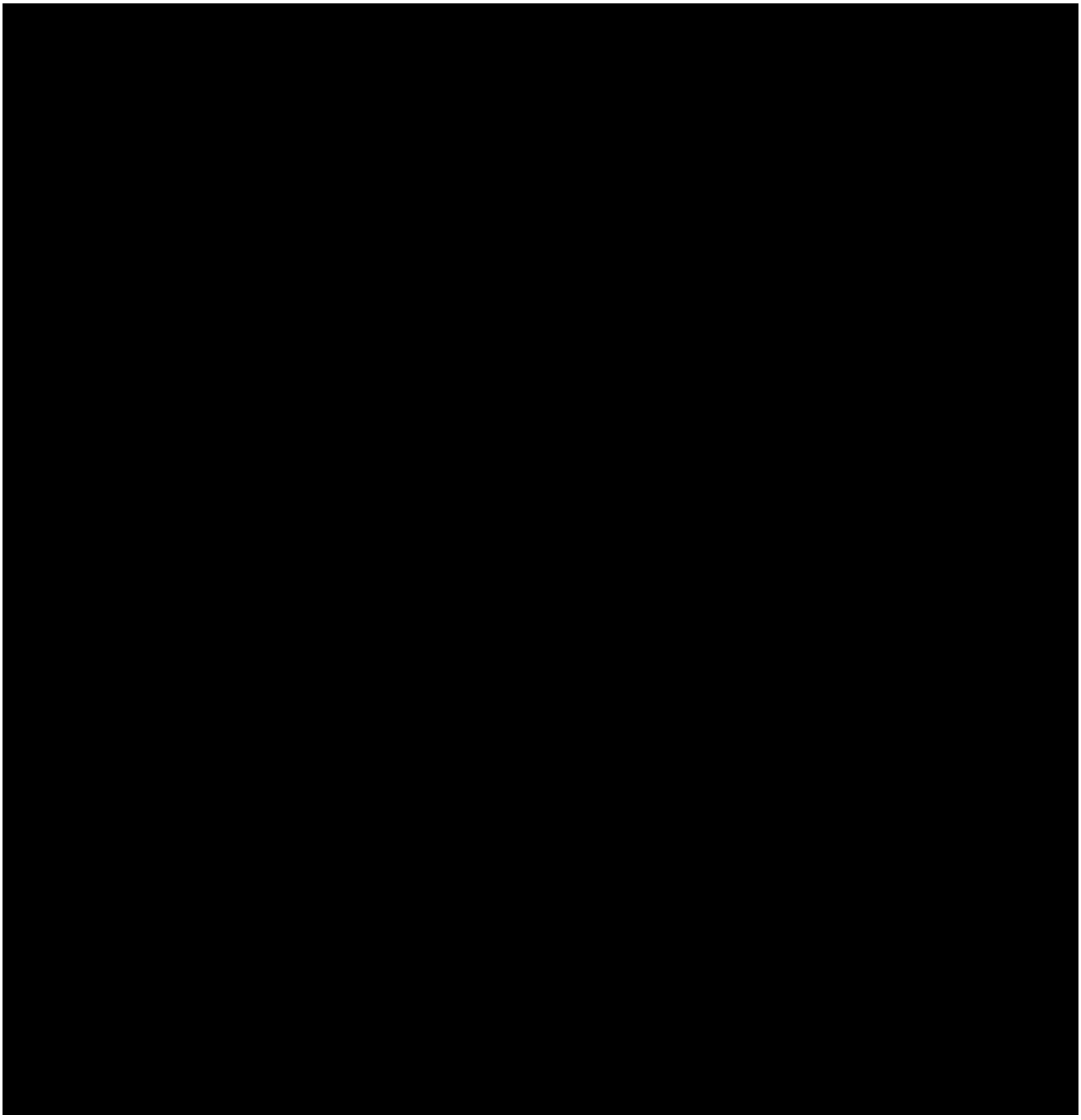
An ECG will be taken at Screening and recorded in source documents. ECG will not be reported in the CSR.

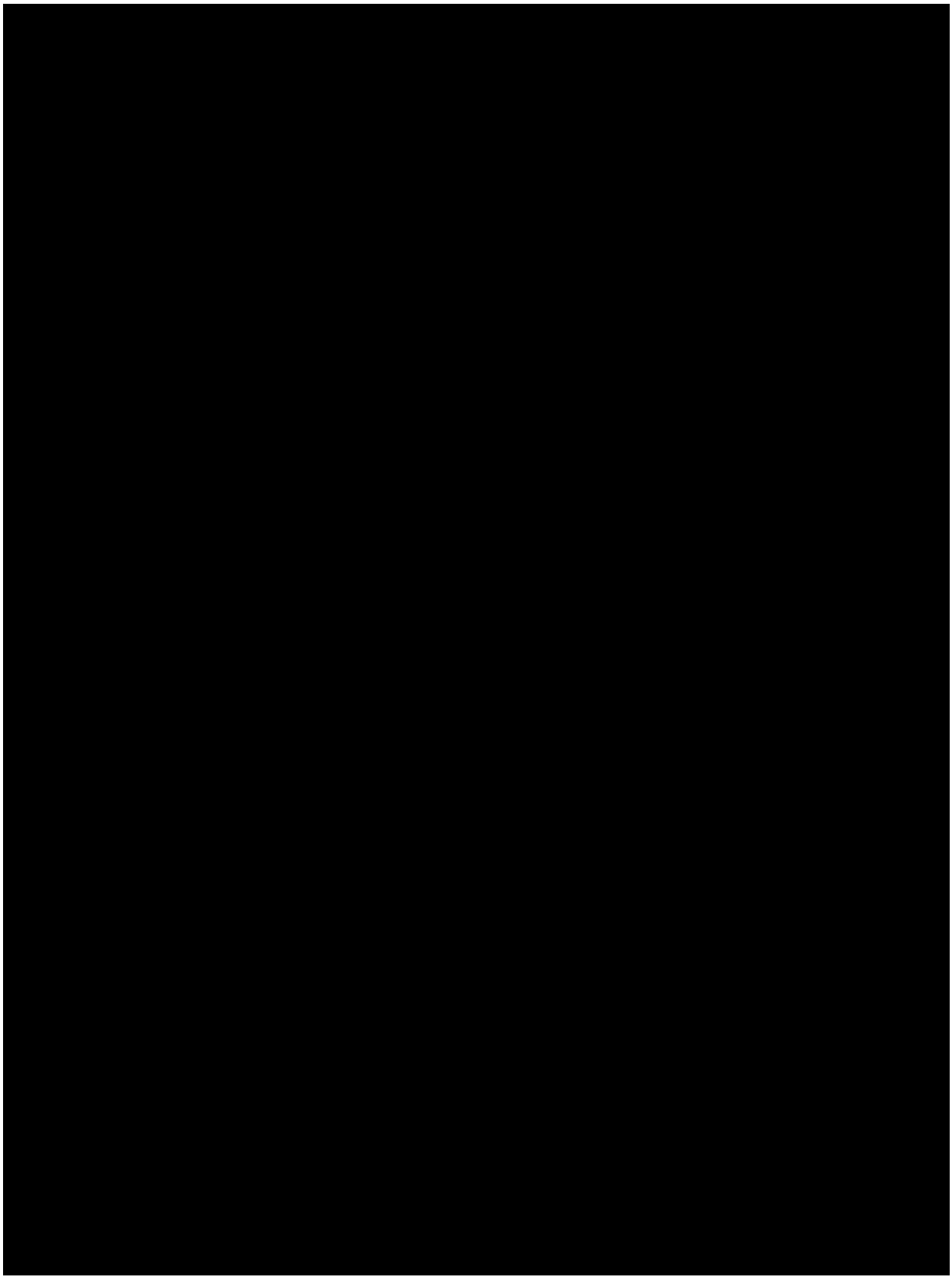
2.8.4.2 Vital signs

Summaries of vital signs data will include treatment-emergent measurements, which are defined as measurements taken post-baseline and up to the last study visit.

Absolute values and change from baseline (where applicable) will be summarized for vital sign parameters (height, weight, temperature, pulse rate, systolic blood pressure, diastolic blood pressure, and respiratory rate) by treatment group, visit and time point including the minimum and maximum treatment-emergent value.

The baseline value is the last value prior to first dose of double-blind treatment.





2.11 Interim analysis

No interim analysis is planned. A DMC at Novartis will be established to conduct periodic semi-blinded or unblinded safety reviews to monitor safety data. The DMC analysis does not inflate the Type I error for the primary efficacy hypothesis testing. Thus no adjustment for multiplicity is required.

More details will be outlined in the DMC charter and a separate DMC SAP.

3 Sample size calculation

Assuming a true treatment difference in proportion of patients meeting the primary composite endpoint of 15% for ruxolitinib added to SoC therapy compared to placebo added to SoC therapy, a sample size of approximately 402 participants provides at least 80% power that the primary analysis will be statistically significant at the two-sided 5% significance level. This holds for a variety of different assumptions with respect to the proportion of patients meeting the primary composite endpoint (Table 3-1).

These sample size assumptions were evaluated in East Version 6.4.

Table 3-1 Sensitivity of sample size assumptions to different proportion of patients meeting the primary composite endpoint for level of significance alpha = 0.05

Proportion of patients meeting the primary composite endpoint in placebo + SoC (%)	Proportion of patients meeting the primary composite endpoint in ruxolitinib + SoC (%)	Power (%)
80	65	91
70	55	85
60	45	82
50	35	82
40	25	85
30	15	91

4 Change to protocol specified analyses

The analysis of the primary endpoint is the same as in the protocol. However, the handling of missing data for the primary composite endpoint is worded differently from the protocol. As the protocol allows patients to be treated for 14 or 28 days some of the patients will not be treated for the entire study period of 29 days because they have recovered and/or been discharged or the investigator feels they would not benefit from further double-blind treatment. Such cases did not "discontinue from the double-blind treatment", but they actually "completed the planned treatment". The off-treatment measurements collected should be used in the analysis (which did not change from the protocol) wherever available, as the outcome is still valid as a result of the previous treatment. In addition, the off-treatment data collected for other patients will also be used, from a intention-to-treatment perspective. The rationale of the missing data handling is made clearer in the SAP.

In Section 12.3 of the protocol, it was mentioned that significant non-drug therapies will summarized according to the ATC classification system. However in the SAP, we specify that surgical and medical procedures will be summarized by primary SOC and PT, as ATC classification is not available for non-drug therapies.

In Section 12.4.2 of the protocol, it was mentioned that baseline clinical status based on the 9-point scale ($\leq 4, \geq 5$) will be included as a covriate in the primary analysis model. In the SAP, we have changed the categorization of the score to be ($\leq 3, \geq 4$). This is becauase very limited patients were recruited with baseline clinical status of 5 and the potential issues it may cause in the modeling.

In Section 12.5.1 of the protocol, it was mentioned that country will be used as a covariate. However, we have modified in the SAP that region will be included instead. This will avoid potential modeling issues due to small sample size in some of the countries.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

Missing/partial start date or end date of double-blind treatment will not be imputed.

5.1.2 AE date imputation

Partial AE start and end dates will be imputed. If there is uncertainty whether an AE occurred treatment-emergent or not, imputation will be performed, such that AE will be considered as treatment-emergent. Rules for imputing AE end date or start date will be provided in Programming Dataset Specification (PDS) document in details.

5.1.3 Concomitant medication date imputation

Rules for imputing the CM end date or start date will be provided in PDS document in details.

5.2 AEs coding/grading

The MedDRA version which will be available at the time of database lock, will be used for the coding purpose of the adverse events.

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per Common Terminology Criteria for Adverse Events (CTCAE) version 5.0.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. CTCAE Grade 5 is not defined for laboratory values.

5.4 Statistical models

5.4.1 Primary analysis

The null hypothesis for the primary analysis is that there is no difference in the proportion of patients meeting the primary composite endpoint between ruxolitinib + SoC and placebo + SoC treatment groups. The alternative hypothesis is that there is a difference between the two treatment groups.

The null hypothesis will be tested using a logistic regression model with treatment, age, gender, region, and baseline clinical status based on the 9-point ordinal scale (≤ 3, ≥ 4) as covariates.

The SAS procedure LOGISTIC will be used to conduct the analysis.

The estimated odds ratio, associated two-sided 95% confidence interval, and two-sided p-value will be presented for treatment contrasts between ruxolitinib + SoC and placebo + SoC.

In case of separability, Firth's penalized maximum likelihood estimation will be performed. The SAS macro %fl (Heinze and Ploner 2004) will be used for performing the estimation.

- **Sensitivity analysis of the primary endpoint**

Non-parametric analysis

In order to determine the robustness of the logistic regression model used for the primary analysis, nonparameteric randomization-based analysis of covariance model (Koch et al 1998) will be evaluated using the same covariates as in the primary analaysis model.

A SAS macro (%NParCov4) has been developed to facilitate the implementation of this analysis (Dmitrienko and Koch 2017).

To test the treatment effect, the most appropriate variance matrix is the one that applies to the randomization distribution under the null hypothesis of no treatment differences: a first model will be obtained using this assumption (i.e. parameter HYPOTH = NULL).

It is also of interest to obtain CI for the odds ratio. The most appropriate variance matrix in this case is the one that applies under the alternative: a second model will be obtained using this assumption (i.e. parameter HYPOTH = ALT).

Risk difference based on the marginal standardization method

The sensitivity analysis will assess the between group comparison based on the risk difference using the marginal standardization method. The estimated risk difference can be derived from a logistic regression model fit to the binary outcome variable with the same covariates as in the primary analysis model. The fitted logistic regression model is used to predict the response risk for every subject in the study as if they had received ruxolitinib + SoC or placebo + SoC, and the difference in risks between treatment groups is computed. The delta method is used to calculate a standard error for the difference.

The method will be implemented using SAS macro (%margins) or the SAS macro provided in the paper by Ge et al 2011.

- **Multiple imputation for supplementary analysis**

For supplemental analysis the following imputation steps need to be performed for missing data. Missing data of the composite endpoint will be imputed using J2R assumption for the ruxolitinib + SoC arm and MAR assumption for placebo arm.

1. Impute missing data:
   Select all patients, impute missing values using the MI approach based on the logistic regression method for 100 time and obtain 100 imputed dataset. A pattern-mixture model approach (Ratitch and O'Kelly 2011) that uses a control-based pattern imputation will be implemented. That is, an imputation model for the missing observations in the ruxolitinib + SoC group is constructed not from the observed data in the ruxolitinib + SoC group but rather from the observed data int the placebo + SoC group. This model is also the imputation model that is used to impute missing observations in the placebo + SoC group. The imputation model will include age, gender, region, BMI (($\leq$ 30.0 kg/m^2, > 30.0 kg/m^2), and baseline clinical status (assessed by the 9-point ordinal scale) as predictors. If any of these predictors cause separability issues for the logistic regression, the levels of the predictors will be combined accordingly.
   This results in 100 imputed datasets.

2. The primary composite endpoint will be analyzed (by imputed dataset) using the final multiply-imputed dataset where all missing values are filled, using a logistic regression model with treatment, age, gender, region, and baseline clinical status (assessed by the 9-point ordinal scale) as predictors.
3. The results for the treatment effect from the 100 datasets will then be combined using Rubin's rule.

5.4.2 Key secondary analysis

There is no key secondary analysis defined in the protocol.

5.5 Rule of exclusion criteria of analysis sets

The following table provides the protocol deviations (PD) and other criteria leading to partial or complete exclusion from the analyses sets.

Deviation ID	Description of Deviation
Deviations leading to exclusion from RAS and Safety set	
INCL01B	Patient or guardian / health proxy did not sign informed consent

5.6 Type of pulmonary/ventilatory support

Table 5-1 Type of pulmonary/Ventilatory support based on the level of oxygen requirement from low to high

Type of support	Require supplemental oxygen	Require (invasive) mechanical ventilation
Low flow nasal oxygen	Yes	No
High flow nasal oxygen	Yes	No
Oxgen via face mask	Yes	No
Non-invasive ventilation	Yes	No
Mechanical ventilation	Yes	Yes
Intubation	Yes	Yes
Tracheostomy	Yes	Yes
ECMO	Yes	Yes

6 References

Available upon request

Cao, B, Wang, Y, Wen, W., et al A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19 (2020) New England Journal of Medicine DOI: 10.1056/NEJMoa2001282 Published 18-Mar-2020 and updated 20-Mar-2020.

Dmitrienko, A. and Koch, G.G., (2017). Analysis of clinical trials using SAS: a practical guide. SAS Institute.

Fine, J.P. and Gray, R.J., (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association, 94(446), pp.496-509.

Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika, 80(1), pp.27-38.

Ge, M., Durham, L.K., Meyer, R.D., Xie, W. and Thomas, N., 2011. Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. Drug information journal: DIJ/Drug Information Association, 45(4), pp.481-493.

Heinze, G. and Ploner, M., (2004). A SAS macro, S-PLUS library and R package to perform logistic regression without convergence problems. Medical University of Vienna, Vienna.

Heinze, G. and Schemper, M., (2002). A solution to the problem of separation in logistic regression. Statistics in medicine, 21(16), pp.2409-2419.

Kalbfleisch, J.D. and Prentice, R.L., (2011). The statistical analysis of failure time data (Vol. 360). John Wiley & Sons.

Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), pp.457-481.

Koch, G.G., Tangen, C.M., Jung, J.W. and Amara I.A., (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine, 17(15-16), pp.1863-1892.

Little, R. and Yau, L., (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. Biometrics, pp.1324-1333.

Molenberghs, G. and Kenward, M., (2007). Missing data in clinical studies (Vol. 61). John Wiley & Sons.

Ratitch, B. and O'Kelly, M., (2011). Implementation of pattern-mixture models using standard SAS/STAT procedures. Proceedings of PharmaSUG.

WHO (2020). A Multi-center, Adaptive, Randomized Blinded Controlled Trial of the Safety and Efficacy of Investigational Therapeutics for the Treatment of COVID-19

WHO (18-Feb 2020). WHO R&D Blueprint. Novel coronavirus. COVID-19 Therapeutic Trial Synopsis (Dated 18-Feb-2020).

Williams, B., Alberti, G., Ball, C., Ball, D., Binks, R. and Durham, L., (2012). Royal College of Physicians, National Early Warning Score (NEWS2), Standardising the assessment of acute-illness severity in the NHS, London.

World Health Organization, (2020). Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance, 28 January 2020 (No. WHO / nCoV / Clinical / 2020.3). World Health Organization.