Statistical Analysis Plan for Study M20-310 (PCYC-1150-IM)

ibrutiNib in SARS CoV-2 induced Pulmonary Injury and Respiratory failure (iNSPIRE)

Date: 20 April 2021

Version 3.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Ibrutinib Study M20-310 titled "ibrutiNib in SARS CoV-2 induced Pulmonary Injury and Respiratory failure (iNSPIRE)."

Study M20-310 examines the efficacy, safety, and tolerability of ibrutinib as an adjuvant therapy to supportive care in hospitalized subjects who presented with coronavirus disease 2019 (COVID-19) related pulmonary distress requiring supplemental oxygen. The exploratory analyses of inflammatory and chemo-attractant cytokines, IgM and IgG antibody titers to SARS-CoV-2 will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

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

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The primary objective of this study will be:

- To evaluate the proportion of subjects alive and without respiratory failure through Day 28 in hospitalized subjects who present with COVID-19 related pulmonary distress requiring supplemental oxygen.

The secondary objectives of this study will be:

- To determine if the addition of ibrutinib to supportive care reduces necessity for hospitalization, length of need for supplemental oxygen, or mechanical ventilation in hospitalized subjects who present with COVID-19 related pulmonary distress requiring supplemental oxygen.

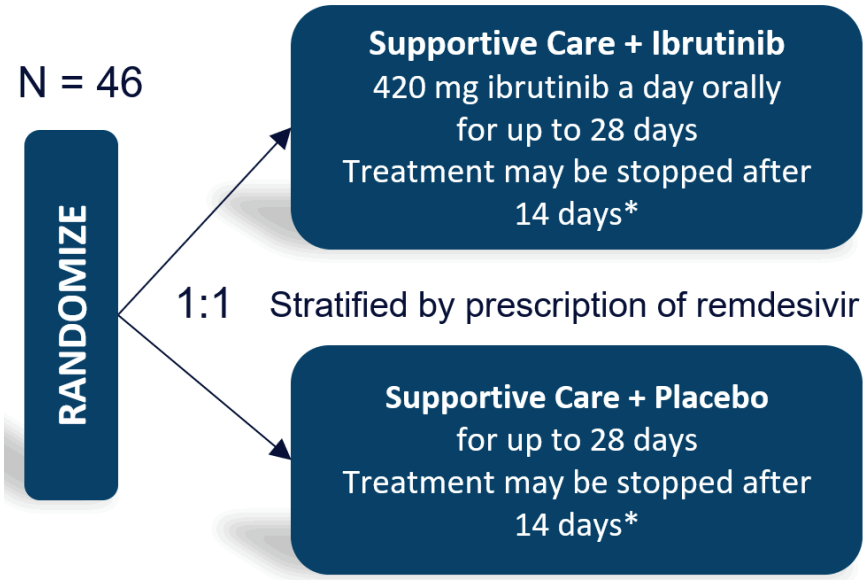
- To assess the safety and tolerability of ibrutinib as an adjuvant therapy to supportive care in hospitalized subjects who present with COVID-19 related pulmonary distress requiring supplemental oxygen.

2.2 Study Design Overview

M20-310 is a multi-center, phase 2, double-blind, placebo-controlled study where approximately 46 subjects will be enrolled and randomized with 1:1 ratio to receive either supportive care and placebo or supportive care and ibrutinib for up to 28 days, with possible replacement for subjects who withdraw consent or are lost to follow up before Day 28. All subjects will be followed to Day 58 with diary and telehealth visits permitted after discharge from hospital.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



* Treatment may be stopped at the discretion of the treating physician after 14 days if the patient is clinically stable, with negative COVID-19 viral nasopharyngeal swab by RT-PCR and has been off supplemental oxygen for > 48 hours.

2.3 Treatment Assignment and Blinding

Subjects meeting eligibility criteria will be randomized to ibrutinib + supportive care vs. placebo + supportive care in a 1:1 ratio. Randomization will be stratified by prescription for remdesivir at baseline.

All subjects will be assigned a unique identification number at the initial screening visit. For subjects who rescreen, they should retain the same identification number at the initial screening visit.

An unblinded AbbVie Central Randomization Person (ACRP) will perform the role of the traditional IRT system by collaborating with the unblinded pharmacists at sites to assign a unique randomization number and the associated treatment (ibrutinib or placebo for ibrutinib) to subjects at baseline based on the computer generated randomization schedules generated by the statistics department at AbbVie. Subjects that withdraw consent or are lost to follow up from the study before Day 28 may be replaced. All randomized subjects who received at least 1 dose of study drug will be included in the Full Analysis Set.

The investigator, study site personnel (other than the unblinded pharmacist) and the subjects will remain blinded to each subject's treatment throughout the study. The sponsor will remain blinded until the primary analysis, i.e., after all subjects have completed the Day 28 Visit or prematurely discontinue the study.

Should a subject withdraw consent during study treatment or is lost to follow up during study treatment, the site will notify the ACRP immediately to allow for possible replacement.

2.4 Sample Size Determination

Forty-six subjects will be randomized at an allocation of 1:1 to supportive care or supportive care plus ibrutinib, stratified by prescription for remdesivir. The power calculations are based on the proportion of subjects alive, without respiratory failure at

28 days in the control arm and in the experimental arm. This sample size will provide at least 80% power to detect the difference between a placebo response rate of 60% and an active response rate of 90% at 1-sided significance level of 0.1, based on a chi-square test. The response rate in the control arm is based on the previously reported clinical outcomes among patients with COVID-19 during hospitalization.[1] No dropout rate is accounted for since subjects that withdraw consent or are lost to follow up from the study before Day 28 may be replaced.

A blinded sample size re-estimation (BSSR) will be conducted when approximately 70% of the planned subjects (i.e., 32 subjects) have completed the Day 28 visit to assess if the sample size estimation assumption holds. If the overall observed response rate is closer to 0.5, sample size increase may be needed. Detailed methodology for the BSSR is described in Section 11.0.

3.0 Endpoints

3.1 Primary Endpoint(s)

The primary endpoint will be the proportion of subjects alive and without respiratory failure through Study Day 28, where respiratory failure is defined as a clinical diagnosis of respiratory failure AND initiation of one of the following therapies:

- endotracheal intubation and mechanical ventilation,
- extracorporeal membrane oxygenation,
- high flow nasal cannula oxygen delivery (i.e., reinforced nasal cannula delivering heated, humidified oxygen with fraction of delivered oxygen ≥ 0.5 and flow rates of ≥ 30 L/min),
- non-invasive positive pressure ventilation,
- clinical diagnosis of respiratory failure with initiation of none of these measures only when clinical decision-making is driven solely by resource limitation.

3.2 Secondary Endpoint(s)

The secondary endpoints include the following:

- Change in the World Health Organization (WHO)-8 point ordinal scale from baseline between the experimental and control arms at Study Day 14
- Median reduction in days spent on supplemental oxygen
- All-cause mortality by Study Days 7, 14, 21, and 28
- Proportion of subjects experiencing respiratory failure or death by Study Days 7, 14, 21, and 28
- Mechanical ventilation-free survival
- Days on mechanical ventilation
- Duration of hospitalization
- Time to discharge
- PaO2:FiO2 and/or oxygenation index

3.3 Other Efficacy Endpoint(s)

There are no additional efficacy endpoints for this study.

3.4 Safety Endpoint(s)

To assess the safety and tolerability of ibrutinib as an adjuvant therapy to standard of care, the following endpoints will be included in the safety analyses:

- Treatment-emergent adverse events
- Serious adverse events
- Adverse events leading to study drug discontinuation
- Other Safety Observations as defined in Section 9.2.5
- Clinical laboratory tests
- 12-lead electrocardiogram (ECG)

3.5 Additional Endpoint(s)

The exploratory analyses of inflammatory and chemo-attractant cytokines, IgM and IgG antibody titers to SARS-CoV-2 will not be covered in this SAP.

4.0 Analysis Populations

The following population sets will be used for the analyses.

Full Analysis Set

The Full Analysis Set (FAS) includes all randomized subjects who received at least 1 dose of study drug. The FAS will be used for all baseline demographics and disease characteristics and efficacy analyses. Subjects will be included in the analysis according to the treatment groups to which they are randomized.

Modified FAS

The modified FAS includes all randomized and dosed subjects who have assessment of the primary endpoint through Day 28 or die before Day 28. A supplementary analysis of the primary endpoint as well as the first 3 secondary endpoints (i.e., change in WHO-8 point ordinal scale from baseline, median reduction in days spent on supplemental oxygen, and all-cause mortality) will be performed on the modified FAS population.

Safety Population

Safety analyses will be performed using the Safety Population which consists of all subjects who received at least 1 dose of study drug. Subjects will be included in the analysis according to the study drug that they actually received. If a subject takes more than 1 treatment, the subject will be grouped in the treatment group for which they received the most doses. Note for the interim analysis, subjects will be grouped according to treatment as randomized since treatment information will not be available until study unblinding.

5.0 Subject Disposition

The total number of subjects who were screened, randomized, and treated will be summarized. Reasons for exclusion, including screen failure, will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group:

- Subjects randomized in the study;
- Subjects who took at least one dose of study drug;
- Subjects who prematurely discontinued study drug;
- Subjects who completed protocol-specified treatment of 28 days;
- Subjects who prematurely discontinued the study;
- Subjects who completed the study;
- Subjects in each analysis population.

The number and percentage of subjects who discontinued study drug by reason (all reasons and primary reason, including discontinuation due to being clinically stable and off of supplemental oxygen for > 48 hours after Day 14) will be summarized overall and by treatment group for the FAS. Similar summaries will be provided for discontinuations from the study.

6.0 Study Drug Duration, Compliance and Dose Intensity

For the Safety Population, duration of treatment will be summarized for each treatment group. Duration of treatment is defined for each subject as last dose date minus first dose date + 1 day. Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum. In addition, the number and percentage of subjects in each treatment duration interval (≤ 14 days, > 14 days) will be summarized.

The number and percentage of subjects will also be summarized for:

- Subjects with reported study drug temporary interruptions by reason as captured on the eCRF;
- Subjects with reported study drug dose reduction by reason as captured on the eCRF.

Treatment compliance is defined as the number of tablets actually taken divided by the number of tablets that should have been taken.

Compliance will be calculated for each subject and summarized by treatment group with the number of subjects treated, mean, median, SD, minimum, and maximum. Percent compliance will be summarized.

Dose intensity is defined for each subject as average daily dose/intended daily dose (i.e., 420 mg), where the average daily dose is defined as the total cumulative dose received (mg) divided by duration of treatment in days. Dose intensity will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum.

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

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the FAS overall and by treatment group. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

The following demographic and characteristics parameters will be summarized.

Subject Demographics

- Sex (male, female)
- Age (years)
- Race (White, Black, American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Multi, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Body weight (kg) at baseline
- Height (cm) at baseline
- BMI (kg/m^2) at baseline

Disease Characteristics

- Stratification factor of prescription for remdesivir (yes, no)
- WHO-8 ordinal scale at baseline (4, 5)
- WHO-8 ordinal scale at screening (4)
- For cirrhotic subjects only: Child-Pugh classification at baseline (not done, A, B)
- SARS-CoV-2 RNA detection at baseline (negative, positive, indeterminate)
- PaO2:FiO2 at baseline (continuous; <100 mmHg, ≥100 mmHg and <200 mmHg, ≥200 mmHg and <300 mmHg, ≥300 mmHg)
- Supplemental oxygen use at baseline (yes, no)
- For subjects on supplemental oxygen only: interface of supplemental oxygen, fraction of oxygen and flow rate (L/min)
- Non-invasive positive pressure ventilation use at baseline (yes, no)
- COVID-19 symptom onset duration prior to baseline (continuous; ≤ median, > median; ≤ 10 days, > 10 days)
- COVID-19 diagnosis duration prior to baseline
- Tobacco user (current, former, never, unknown)
- Alcohol status (current, former, never, unknown)

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

7.3 Prior and Concomitant Medications

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

8.0 Efficacy Analyses

8.1 General Considerations

All efficacy analyses will be conducted for the FAS unless otherwise specified. All efficacy tests will be 2-sided at an alpha level of 0.2.

The Primary Analysis will be performed after all ongoing subjects have completed the Day 28 visit or have prematurely discontinued from study and the database has been locked.

"Baseline" refers to the last non-missing observation on or before the first day of study drug administration.

8.2 Handling of Missing Data

Missing data could occur due to various reasons, including missed assessments, early dropout from the study, etc. In general, if a subject takes a BTK inhibitor besides the study drug, then all data after the start of that BTK inhibitor will be set to missing. Handling of missing data will be described for the different types of efficacy endpoints as follows.

Primary Endpoint of Being Alive and Without Respiratory Failure

If a subject is missing daily records on the Supplemental Oxygen or Invasive Ventilation - Mechanical/Cardiopulmonary Support eCRFs before Day 28, but has an WHO-8 point ordinal score recorded to fill in those days then it will be used to record respiratory failure or not (i.e., an ordinal score of 4 or below indicates no respiratory failure or death on the same day).

After the above, subjects with no record of death or respiratory failure and one of the following events will be analyzed as follows:

- If a subject has missing responses on days after being discharged from hospital prior to Day 28 as indicated on the Hospitalization eCRF without any re-hospitalization record, the subject will be categorized as a success.
- If a subject has missing responses on days after completing study drug with a reason of "clinically stable and off of supplemental oxygen for > 48 hours after Day 14" as indicated on the study drug completion eCRF without any re-hospitalization record, the subject will be categorized as a success.
- If a subject takes a BTK inhibitor besides the study drug (and does not meet either condition in the previous 2 bullets), then all data after the start of that BTK inhibitor will be set to missing.

If a subject has missing responses of death or respiratory failure for ≤ 2 consecutive days and is a success for both the preceding response and succeeding response, the missing responses on the day(s) will be imputed as a success.

If a subject has missing responses for > 2 successive days or monotone missing data before Day 28 without WHO-8 point ordinal score to fill in those days, the missing responses will be handled with multiple imputation method with fully conditional specification (FCS) approach to generate 30 datasets using the logistic regression method. The imputation model will include treatment, stratification factor of prescription for remdesivir, and duration of symptoms before Day 1 (continuous). The random seed for PROC MI will be the SAS numerical form of the first subject randomization date in the study. In case of non-convergence, the random seed will be updated by adding 100000 at each attempt until convergence of model occurs. If the model will not converge due to data sparsity issue, then stratification factor will be removed from the imputation model. Using the Miettinen and Nurminen (MN) method[2] for the difference in the response rates between treatment groups adjusting for the stratification factor of prescription for remdesivir, the imputed endpoints will be analyzed using each of the 30 datasets. PROC MIANALYZE will be used to generate the final inferences of the risk difference between treatment groups.

WHO Ordinal Scale

The secondary endpoint of change in the WHO-8 point ordinal scale from baseline between treatment groups at Day 14 will be determined based on subjects' clinical status as indicated on the Ordinal Scale for Clinical Improvement eCRF at both Day 1 and the Day 14 visit. Subjects without records of ordinal scale on Day 1 will be excluded from analysis.

If a subject takes a BTK inhibitor besides the study drug, then all data after the start of that BTK inhibitor will be set to missing.

Subjects with no record of ordinal scale at the Day 14 visit due to the following events and without oxygen therapy recorded to fill in the days will be analyzed as follows:

- If a subject has been discharged from hospital without record of re-hospitalization or death before the Day 14 visit, then the missing ordinal score for the Day 14 visit will be imputed to the minimum of 2 or their last non-missing score before the Day 14 visit window.
- If a subject died before reaching the Day 14 visit, then the missing ordinal score for the Day 14 visit will be imputed as 8.

Similar imputation will be performed for subjects with no records of ordinal scale at the Day 7 visit and Day 28 visit.

Days on Supplemental Oxygen, Mechanical Ventilation and of Hospitalization

For the analysis of the days spent on supplemental oxygen, days on mechanical ventilation, or duration of hospitalization, the number of days will be imputed to the maximum number of days on study drug (28) following the death of a subject or if data is missing (e.g., due to starting another BTK inhibitor during the time period).

If a subject prematurely discontinued from study without the date of the subject being off the supplemental oxygen then the subject's premature discontinuation date will be used for the duration calculation. Similarly, if a subject prematurely discontinued from study without the date of the subject being off mechanical ventilation, or if a subject prematurely discontinued from study without a discharge date, then the subject's premature discontinuation date will be used for the duration calculation.

All-Cause Mortality

All deaths through Day 28 will count in all-cause mortality, even deaths after starting another BTK inhibitor.

PaO2:FiO2 and oxygenation index

For continuous endpoints of PaO2:FiO2 or oxygenation index, missing data will be handled using a Mixed-Effect Model Repeated Measurement (MMRM) model.

- The MMRM will be conducted using a mixed model including observed measurements at Study Days 7, 14, 21 and 28. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, stratification factor for prescription of remdesivir, and the continuous fixed covariate of Baseline measurement. An unstructured variance covariance matrix (UN) will be used. If the model cannot converge, an appropriate covariance structure matrix (e.g., autoregressive (1) or compound symmetry) will be used. The parameter estimations are based on the method of restrictive maximum likelihood (REML). The fixed effects will be used to report model-based means at corresponding visits.

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary endpoint will be the proportion of subjects alive and without respiratory failure through Study Day 28, where respiratory failure is defined as a clinical diagnosis of respiratory failure AND initiation of one of the following therapies:

- endotracheal intubation and mechanical ventilation,
- extracorporeal membrane oxygenation,
- high flow nasal cannula oxygen delivery (i.e., reinforced nasal cannula delivering heated, humidified oxygen with fraction of delivered oxygen ≥ 0.5 and flow rates of ≥ 30 L/min),
- non-invasive positive pressure ventilation,
- clinical diagnosis of respiratory failure with initiation of none of these measures only when clinical decision-making is driven solely by resource limitation.

For the primary efficacy endpoint, a subject will be defined as a failure if the subject died or had a clinical diagnosis of respiratory failure with initiation of any of the 5 therapies as indicated on the "Invasive ventilation – Mechanical/Cardiopulmonary support" and/or "Supplemental Oxygen" eCRFs on any single day from Day 1 through Day 28. If the subject had no clinical diagnosis of respiratory failure (or had a clinical diagnosis of respiratory failure without fulfilling any of the bullets above) and was alive from Day 1 through Day 28, the subject will be defined as a success. The days of the clinical diagnosis and respiratory therapies need not coincide.

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint

The handling of missing data for the primary efficacy endpoint is described in Section 8.2.

8.3.3 Primary Efficacy Analysis

The number and percentage of subjects alive and without respiratory failure will be summarized by treatment arm, within each stratum and across strata. The difference in response rates between the experimental arm and the control arm will be analyzed using the Miettinen–Nurminen (MN) method adjusting for the stratification factor of prescription for remdesivir. MN test p-value for testing the rate difference = 0 at two-sided alpha = 0.2 and associated 80% and 95% MN confidence intervals for the rate difference will be reported.

If there is a stratum for a treatment arm that has 0 subjects in any cell in the contingency table, the MN test will be performed by combining the strata or the Cochran-Mantel-Haenszel (CMH) test will be used if needed.

8.3.4 Additional Analyses of the Primary Efficacy Endpoint

A supplementary analysis of the primary endpoint will be performed on the FAS population excluding subjects who developed respiratory failure before the first dose of study drug administration.

In addition, a supplementary analysis of the primary endpoint as well as the first 3 secondary endpoints (i.e., change in WHO-8 point ordinal scale from baseline, median reduction in days spent on supplemental oxygen, and all-cause mortality) will be performed on the modified FAS population. A supplementary analysis of the primary endpoint will be performed on the modified FAS populations using the MN method as described above for the primary analysis.

8.4 Secondary Efficacy Analyses

The following secondary efficacy endpoints will be analyzed by treatment group within each stratum and across strata.

- Change in the World Health Organization (WHO)-8 point ordinal scale from baseline between the experimental and control arms at Study Day 14
- Median reduction in days spent on supplemental oxygen
- All-cause mortality by Study Days 7, 14, 21, and 28
- Proportion of subjects experiencing respiratory failure or death by Study Days 7, 14, 21, and 28
- Mechanical ventilation-free survival
- Days on mechanical ventilation
- Duration of hospitalization
- Time to discharge
- PaO2:FiO2 and/or oxygenation index

To evaluate the odds of clinical status improvement as determined by the WHO-8 point ordinal scale (Appendix C) from baseline at the Day 14 visit between treatment groups, a proportional odds model will be conducted using the ordinal score change from baseline to the Day 14 visit as the dependent variable, with treatment as the main effect, and the stratification factor of prescription for remdesivir and baseline ordinal score as covariates. If model does not converge, the baseline ordinal score and the stratification factor will be removed one at a time to check if the model will converge. If quasi-complete separation issue occurs, then the separated ordinal levels may be combined as appropriate. The point

estimate for the odds ratio between experimental arm and control arm and the associated 80% and 95% Wald confidence intervals will be provided, as well as the p-value for testing the odds ratio = 1 from the Wald chi-square test. A p-value from the score test for the proportionality assumption will also be provided. If the score test fails to satisfy the proportionality assumption, then a multinomial logistic regression model may be used.

Similar analyses will be used to analyze the odds of clinical status improvement as determined by the WHO-8 point ordinal scale from baseline to Day 7 visit and Day 28 visit between treatment groups.

In addition, proportional odds models will be used to analyze the WHO ordinal scale scores at Day 7 visit, Day 14 visit and Day 28 visit, respectively.

The stacked bar plots will be provided both for proportion of subjects in each score of the ordinal scale and each category of change from baseline in ordinal scale score by treatment arm at the Day 7 visit, Day 14 visit and Day 28 visit.

A supplementary analysis of the change of WHO-8 point ordinal scale from baseline at the Day 14 visit will be performed by using the WHO-8 point ordinal scale at screening as baseline for all subjects.

In addition, the number and proportion of subjects with at least 2 points of clinical improvement as indicated by the WHO ordinal scale from baseline to the Day 14 visit will be summarized by treatment arm, within each stratum and across strata. The MN test p-value for testing the rate difference = 0 at two-sided alpha = 0.2 and associated 80% and 95% MN CIs for the rate difference across strata will also be presented. The 80% and 95% MN CIs for the rate difference within each stratum will also be presented. If there is a stratum for a treatment arm that has 0 subjects in any cell in the contingency table, the MN test will be performed by combining the strata or the CMH test will be used if needed. A supplementary analysis of the proportion of subjects with at least 3 points of clinical improvement as indicated by the WHO ordinal scale from baseline to the Day 14 visit will also be performed using the same procedures.

For the analysis of median reduction in days spent on supplemental oxygen, mechanical ventilation or duration of hospitalization, the duration in number of days are defined as follows:

- Days spent on supplemental oxygen = the date of the subject being off the supplemental oxygen – date of initiation of supplemental oxygen + 1 day.
- Days spent on mechanical ventilation = the date of the subject being off mechanical ventilation – date of initiation of mechanical ventilation + 1 day.
- Duration of hospitalization = the hospitalization discharge date –hospitalization admission date + 1 day.

If a subject received more than 1 period of supplemental oxygen therapy during the study or switched from supplemental oxygen to a more intensive therapy, then the days spent on supplemental oxygen will be calculated as the sum of all the periods where the subject is on supplemental oxygen or a more intensive therapy through Day 28. Similar calculation will be applied to the number of days on mechanical ventilation or duration of hospitalization through Day 28. If the date of the first initiation of supplemental oxygen is before Baseline Day 1, then the days spent on supplemental oxygen will be calculated from Baseline Day 1 to the date of the subject being off the supplemental oxygen. Similar calculation will be applied to duration of hospitalization through Day 28.

Descriptive statistics will be presented for days spent on supplemental oxygen through Day 28. Median days spent on supplemental oxygen will be compared between treatment groups using Van Elteren's test[3] with the stratification factor of prescription for remdesivir. Similar analysis will be performed on the number of days on mechanical ventilation and duration in number of days of hospitalization.

Number and percentage of subjects with all-cause mortality by Day 7, Day 14, Day 21, and Day 28 will be summarized by treatment group, within each stratum and across strata. The MN test p-value for testing the rate difference = 0 at two-sided alpha = 0.2 and associated 80% and 95% MN CIs for the rate difference across strata will be reported, and 80% and 95% MN CI for the rate difference within each stratum will be presented. If

there is a stratum for a treatment arm that has 0 subjects in any cell in the contingency table, the MN test will be performed by combining the strata or CMH test will be used if needed. Similar analysis will be performed on the number and percentage of subjects experiencing respiratory failure by Day 7, Day 14, Day 21, and Day 28.

Mechanical ventilation-free survival is defined as the number of days from Baseline Day 1 to the date when a subject initiated mechanical ventilation or died, whichever occurred first, during the 28 days post baseline. If the specified event does not occur by Day 28, subjects will be censored. Specifically, a subject without any post-baseline assessment record will be censored at Baseline Day 1, a subject who prematurely discontinued from study without a record of death or start of mechanical ventilation will be censored at the earlier timepoint of the date of study discontinuation or Day 28, an on-going subject in the study without a record of death or start of mechanical ventilation will be censored at the earlier timepoint of the date of the last evidence that the subject is not on mechanical ventilation or Day 28.

Time to discharge from hospital is defined as the number of days from Baseline Day 1 to the date of the last evidence that a subject is last discharged from hospital, during the 28 days post baseline. If the specified event does not occur by Day 28, subjects will be censored. Specifically, a subject without any postbaseline assessment record will be censored at Baseline Day 1, a subject who died will have time to discharge from hospital censored at Day 28, a subject who prematurely discontinued from study without a record of hospitalization discharge will be censored at the earlier timepoint of the date of study discontinuation or Day 28, an on-going subject in the study without a record of hospitalization discharge will be censored at the earlier timepoint of the date of the last evidence that the subject is hospitalized or Day 28.

For the analysis of mechanical ventilation-free survival and time to discharge from hospital, respectively, the distribution of time to event will be estimated by treatment group using Kaplan-Meier methodology and compared between the experimental arm and the control arm using the log-rank test stratified by prescription for remdesivir and p-value presented. If there is no event in one of the strata, then the unstratified log-rank test will

be performed. Median event onset time will be calculated and the 80% and 95% confidence intervals will be constructed separately by treatment group.

The restricted mean survival time for the mechanical ventilation-free survival and time to discharge, respectively, will be analyzed by comparing the areas under the survival functions between treatment groups. Mean and standard error for the areas under the survival functions will be provided for each treatment group. The p-value and the 80% and 95% confidence intervals from the between-group-comparison will also be provided.

For the analysis PaO2:FiO2 and oxygenation index, descriptive statistics (number of observations, mean, median, standard deviation, median, minimum and maximum) will be presented for the observed values of PaO2:FiO2 and oxygenation index at Study Days 7, 14, 21 and 28, respectively.

Mean difference in the PaO2:FiO2 values between treatment groups at Study Days 7, 14, 21 and 28 will be compared using a MMRM model as described in Section 8.3.2. Point estimates and 80% and 95% CIs for the mean difference in the PaO2:FiO2 between treatment groups will be provided. A similar MMRM model will be performed to analyze the oxygenation index.

8.5 Efficacy Subgroup Analyses

Subgroup analysis of the primary efficacy endpoint will be performed by baseline age (18 to < 40 yr, 40 to < 65 yr, ≥ 65 yr), race, ethnicity, sex, COVID-19 symptom onset duration (≤ median, > median), baseline ordinal score (4, 5), and use of dexamethasone as well as other emerging treatment(s) for COVID-19, if applicable.

The number and percentage of subjects alive and without respiratory failure and for the difference in response rates between treatment group will be presented for each subgroup. Any subgroup levels with fewer than 10 subjects will be combined with other levels as appropriate for analyses. The 2-sided 80% and 95% confidence intervals of the difference in response rates using MN method will be produced if there are at least 10 subjects in

either of the treatment groups of the subgroup. For any subgroup, if there are 0 subjects within a stratum in any treatment group, the unstratified MN method will be used.

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized for the Safety Population. Safety summaries will be presented by treatment group. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest toxicity grade and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs (TEAEs) are defined as any AE with the onset date that is after the first dose of study drug and no more than 30 days after the last dose of study drug. Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

In addition, a listing of subjects experiencing AE related to study drug according to the investigator prior to first dose of study drug or post 30 days after the last dose of study drug will be provided.

9.2.2 Adverse Event Overview

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

- Any TEAE
- Any TEAE related to study drug according to the investigator
- Any TEAE of grade 3 or higher
- Any serious TEAE
- Any TEAE leading to discontinuation of study drug
- Any TEAE leading to study drug dose reduction
- Any TEAE leading to death
- AEs identified as Other Safety Observations as specified in Appendix A
- All deaths

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

Treatment-emergent adverse events will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum toxicity grade and SOC and PT; and by subject number and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest toxicity grade and level of relationship to investigational product will be reported.

In addition, all TEAEs will be summarized by PT and sorted by decreasing frequency for the experimental arm.

9.2.4 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation or Dose Reduction

SAEs (including deaths) and AEs leading to study drug discontinuation or dose reduction will be summarized by SOC and PT and in listing format.

9.2.5 Other Safety Observations

The following Other Safety Observations will be identified per Standardized MedDRA Queries (SMQs), SOC and/or PT search. Detailed information about the search criteria are provided in Appendix A.

- Hemorrhagic events and major hemorrhage
- Cytopenic AEs
- Infections
- Atrial fibrillation
- Cardiac arrhythmia (including ventricular tachyarrhythmias)
- Hepatic disorders
- Interstitial lung disease
- Hypertension
- Ischemic stroke

Number and percentage of subjects experiencing at least one treatment-emergent AE identified as Other Safety Observations will be summarized and presented by SOC and PT. Tabular listings of Other Safety Observations also will be provided.

9.3 Analysis of Laboratory Data

Laboratory measurements of total bilirubin, AST, ALT, albumin, glucose, creatinine, sodium, potassium, calcium, CPK, GGT, alkaline phosphatase, hemoglobin, platelets, WBC, neutrophils, lymphocytes, and INR will be summarized.

Each laboratory variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median,

minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between treatment groups (experimental arm vs. control arm).

Changes in laboratory parameters will be tabulated using shift tables categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. A shift table from baseline to the minimum and maximum value during treatment will be created. Similar shift table(s) will be provided to summarize shifts from baseline to the final post-baseline value.

The number and percentage of subjects with a maximum of at least CTC Grade 3 during treatment for the laboratory parameters as defined in Appendix B will be summarized. The post-baseline value must be in a toxicity grade that is more extreme than the toxicity grade corresponding to the baseline value in order to be counted. A listing of all relevant laboratory parameters will be provided for each subject who had an increase to Grade 3 or higher will be provided.

The following criteria will be used to assess for potential hepatotoxicity.

- ALT > 3 × ULN, > 5 × ULN, > 10 × ULN, > 20 × ULN
- AST > 3 × ULN, > 5 × ULN, > 10 × ULN, > 20 × ULN
- TBL > 1.5 x ULN, > 2 × ULN
- ALT and/or AST > 3 × ULN and TBL > 1.5 × ULN
- ALT and/or AST > 3 × ULN and TBL > 2 × ULN
- ALT > 3 × ULN and TBL > 1.5 × ULN
- ALT > 3 × ULN and TBL > 2 × ULN
- Alkaline phosphatase > 1.5 × ULN

The number and percentage of subjects with laboratory values meeting the above criteria during treatment will be summarized. Listing of ALT, AST, total bilirubin, and alkaline phosphatase values will be provided for each subject who met one or more of the criteria defined above.

An Evaluation of Drug Induced Serious Hepatotoxicity (eDISH) plot of the maximum post-baseline ALT value (as a multiple of the ULN) vs. the maximum post-baseline total bilirubin value (as a multiple of the ULN) during treatment, not necessarily concurrent, will also be utilized to assess for potential hepatotoxicity.

9.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, body temperature, respiratory rate and oxygen saturation will be summarized.

Each vital sign variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for each vital sign variable, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between treatment groups (experimental arm vs. control arm).

Potentially clinically significant vital sign abnormalities will be assessed by the investigators and reported as adverse events.

9.5 Other Safety Analyses

The number and percentage of subjects in following types of mechanical ventilation at the Day 1, Day 7, Day 14, Day 21 and Day 28 visits will be summarized by treatment group.

- endotracheal
- transtracheal

ECG findings will be summarized by treatment group at baseline and the Day 7, Day 14, Day 21 and Day 28 visit. Summaries will include n (%) of patients in the following categories:

- Normal
- Abnormal – Not Clinically Significant
- Abnormal – Clinically Significant
- Unable To Evaluate
- Missing

10.0 Interim Analyses

An interim analysis will occur when 18 subjects have been randomized and treated/followed until availability of the primary endpoint. A database snapshot will be performed for this analysis. Efficacy assessment will be performed using the FAS and will include a futility analysis of the primary endpoint.

The primary analysis for this study will take place after all subjects have completed the Day 28 visit or discontinue from study prior to Day 28, appropriate data cleaning has been performed and the database has been locked for analysis.

In order to ensure the sample size estimation assumption holds for the study, a blinded sample size re-estimation (BSSR) will be conducted when approximately 70% of the planned subjects (i.e., 32 subjects) have completed the Day 28 visit or have data available for the primary endpoint. The objective of analysis is to re-estimate sample size increase if needed. A study database snapshot will be performed for this analysis. The BSSR analysis will be performed by the AbbVie study statistician on the blinded efficacy data. Only the overall response rate across the treatment groups is used in the BSSR. Therefore, the Type-I error will not be inflated, and no multiplicity adjustment is needed for the BSSR.

Detailed Methodology for BSSR

With an initial sample size of N_0 = 46 subjects, the study is powered at > 80% if the ratio of the response rate for the primary endpoint in the experimental arm to the control arm is 1.5 with the overall response rate of 75% across arms.

For a test of statistical superiority, assuming a ratio of $r = p_t^*/p_c^*$, and p^* is the overall proportion of successes across treatment groups, the sample size N is determined by:[4]

$$N = \frac{[\, z_\alpha\sqrt{2p^*(1-p^*)} + z_\beta\sqrt{p_c^*(1-p_c^*) + p_t^*(1-p_t^*)}]^2}{(p_t^* - p_c^*)^2}$$

where z_α and z_β are the critical z-scores to control the Type-I and Type-II errors at α and β, respectively.

In order to incorporate a cap on the sample size (i.e., $N_{cap} = 2 * N_0 = 92$), the final recalculated sample size will be

$$N_{recal} = \min(\max(N_0, N), N_{cap})$$

Based on the formula above, the final recalculated sample size will always be greater than or equal to the initial sample size of 46 evaluable subjects and less than or equal to the limit of 92 evaluable subjects.

If the final sample size is increased <= 20% from N_0 (i.e., final sample size <= 56 subjects), then the primary analysis will occur after all subjects per the increased sample size have completed the Day 28 visit or discontinued from study prior to Day 28. At that point, the Sponsor will be unblinded.

If the final sample size is increased > 20% from N_0 (i.e., final sample size > 56 subjects), then an additional interim analysis will be performed when 46 subjects have completed the Day 28 or discontinued from the study prior to Day 28 where the study team will remain blinded and a DMC will be utilized to review efficacy results (also see Section 10.1 and Section 11.0).

Interim Database Lock

An interim lock will be performed for the primary analysis when all subjects have completed the Day 28 visit (including those who discontinue from study prior to Day 28). Data for the primary analysis will be locked after data cleaning. Data after the Day 28 visit will be added to a new version of the database which will be cleaned and locked at the end of the study.

If an interim analysis based on 46 subjects' available data for the primary endpoint is to be conducted, an interim database lock will be performed after appropriate data cleaning before this interim analysis will be conducted for DMC review.

10.1 Data Monitoring Committee

An external data monitoring committee (DMC) composed of multiple clinicians and 1 statistician who are independent of AbbVie and with relevant expertise in their field will review and monitor unblinded toxicity and accrual data from Study M20-310 on an ongoing basis. The safety and interim efficacy of the study will be monitored by the DMC as outlined below.

An early safety evaluation will occur by the DMC following the enrollment of approximately 3-6 subjects with at least 7 days of treatment/follow up and will continue in regular frequency approximately every 3 – 4 weeks or when an additional 4-5 subjects have been enrolled (and have at least 7 days of follow-up) since the last DMC meeting, whichever is longer, (or ad-hoc as needed) during study drug treatment until all subjects reach Day 28 or prematurely discontinue from study.

In addition, an interim efficacy assessment will be conducted for DMC review after 18 subjects have been randomized and treated/followed until availability of the primary endpoint. Efficacy assessment will be performed using the FAS and will include primary endpoint and all secondary efficacy endpoints, and a non-binding futility analysis of the primary endpoint (also see Section 11.0).

An additional interim efficacy analysis may be conducted (see Section 10.0 for the triggering condition) for DMC review after 46 subjects have been randomized and treated/followed until availability of the primary endpoint. Efficacy assessment will be performed using the FAS and will include primary endpoint and all secondary efficacy endpoints. A superiority analysis and a futility analysis of the primary endpoint will also be conducted (also see Section 11.0).

The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study and to monitor the efficacy profile of the accumulating data and provide recommendations to AbbVie on whether to continue, modify, or terminate the study. A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

11.0 Overall Type-I Error Control

A group sequential design[5] will be used to provide strong control of the type 1 error rate at one-sided alpha of 0.1 across the interim analysi(e)s and the primary analysis comparing the experimental arm to the control arm with respect to the primary endpoint.

Specifically, for the first interim efficacy analysis conducted for DMC review based on 18 subjects' availability data for the primary endpoint, a futility boundary using the gamma family spending function with a conservative boundary (gamma=-4) will be used for futility analysis. Since no superiority analysis is to be performed, no alpha will be spent at this interim and thus no Type I error inflation is associated with this interim analysis.

If a second interim analysis based on 46 subjects' availability data for the primary endpoint is to be conducted (see Section 10.0 for the triggering condition), the Lan-DeMets alpha spending function based on O'Brien-Fleming boundary will be used to decide a pre-specified weight of alpha allocation for the superiority analysis of the primary endpoint at this interim. A futility analysis using the gamma family spending

function with a conservative boundary (gamma=-4) will also be performed on the primary endpoint at this interim.

12.0 Version History

Table 1. SAP Version History Summary

Version	Date	Summary
1.0	08 July 2020	Original version
2.0	17 August 2020	• Updates were made to add additional efficacy analysis of both the WHO ordinal scale at Day 7, 14 and 28 visits and the change from baseline in scale at each of the visits. • Added analysis of the areas under the survival functions for the mechanical ventilation-free survival and time to discharge. • Added 95% confidence intervals to all efficacy analysis and clarified p-values to be based on 2-sided tests at an alpha level of 0.2.
3.0	20 April 2021	• Updated the DMC meeting frequency based on subject enrollment to ensure sufficient new data are available for each review. • Added additional analysis of the primary endpoint and the secondary endpoint of WHO-8 point ordinal scale from baseline. • Updated handling of missing data for WHO-8 point ordinal scale to incorporate measurements following hospital discharge in imputation. • Added Ordinal Scale for Clinical Improvement as Appendix C

13.0 References

1. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020;323(20):2052-9.
2. Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med. 1985;4(2):213-26.

3. Van Elteren PH. On the combination of independent two-sample tests of Wilcoxon. Bull Int Stat Inst. 1960;37:351-61.
4. Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: a review. J Biopharm Stat. 2005;15(4):559-74.
5. Jennison C, Turnbull BW. Group Sequential Methods with Applications to Clinical Trials. New York: Chapman & Hall; 2000.

Appendix A. Definitions of Other Safety Observations

Other Safety Observations will be identified using the following search criteria:

Other Safety Observations	Search Criteria	Include in AE Overview (Y/N)
Hemorrhagic events and major hemorrhage	• Hemorrhage SMQ excluding laboratory terms (SMQ 20000039 EXCL); narrow search • Sub-analysis of Hemorrhage (SMQ 20000039 EXCL) limited to hemorrhagic events of grade 3+, serious hemorrhagic events, or CNS hemorrhage of any grade	Y
Cytopenic AEs	PTs of "Neutropenia" (PT 10029354), "Febrile neutropenia" (PT 10016288), "Thrombocytopenia" (PT 10043554), and "Anaemia" (PT 10002034)	Y
Infections (including viral reactivation)	• SOC of "Infections and Infestations" (SOC 10021881) • Sub-analysis of Opportunistic infections (SMQ 20000235)	Y
Atrial fibrillation	PT of "Atrial fibrillation" (PT 10003658)	Y
Cardiac arrhythmia (including ventricular tachyarrhythmias)	• Cardiac arrhythmias (SMQ 20000049) excluding PT of "Atrial fibrillation" • Ventricular tachyarrhythmias (SMQ 20000058); narrow search	Y
Hepatic disorders	SOC of "Hepatobiliary Disorders" (SOC 10019805)	Y
Interstitial lung diseases (ILD)	ILD (SMQ 20000042); narrow search	Y
Hypertension	Hypertension (SMQ 20000147); narrow search	Y
Ischaemic stroke	Ischaemic central nervous system vascular conditions (SMQ 20000063); narrow search	Y

Appendix B. Definitions of Toxicity Grades 1, 2, 3 and 4 for Laboratory Values

Test	Grade 1	Grade 2	Grade 3	Grade 4
ALT/SGPT	> ULN – 3 × ULN	> 3 – 5 × ULN	> 5 – 20 × ULN	> 20 × ULN
AST/SGOT	> ULN – 3 × ULN	> 3 – 5 × ULN	> 5 – 20 × ULN	> 20 × ULN
Alkaline Phosphatase	> ULN – 2.5 × ULN	> 2.5 – 5 × ULN	> 5 – 20 × ULN	> 20 × ULN
Total Bilirubin	> ULN – 1.5 × ULN	> 1.5 – 3 × ULN	> 3 – 10 × ULN	> 10 × ULN
Albumin	< LLN – 30 g/L	< 30 – 20 g/L	< 20 g/L	--
Creatinine	> ULN – 1.5 × ULN	> 1.5 – 3 × ULN	> 3 – 6 × ULN	> 6 × ULN
CPK	>ULN - 2.5 × ULN	>2.5 - 5 × ULN	>5 - 10 × ULN	>10 × ULN
GGT	>ULN - 2.5 × ULN	>2.5 - 5 × ULN	>5 - 20 × ULN	>20 × ULN
Hemoglobin	< LLN – 100 g/L	< 100 – 80 g/L	< 80 g/L	--
Neutrophils	< LLN – 1.5 × 10^9/L	< 1.5 – 1.0 × 10^9/L	< 1.0 – 0.5 × 10^9/L	< 0.5 × 10^9/L
Platelets	< LLN – 75.0 × 10^9/L	< 75.0 – 50.0 × 10^9/L	< 50.0 – 25.0 × 10^9/L	< 25.0 × 10^9/L
Lymphocytes	< LLN – 0.8 × 10^9/L	< 0.8 – 0.5 × 10^9/L	< 0.5 – 0.2 × 10^9/L	< 0.2 × 10^9/L
WBC	<LLN - 3.0 × 10^9/L	<3.0 - 2.0 × 10^9/L	<2.0 - 1.0 × 10^9/L	<1.0 × 10^9/L
INR	>1 - 1.5 × ULN	>1.5 - 2.5 × ULN	>2.5 × ULN	--

Note: Definitions based on NCI-CTCAE Version 4.03.

Appendix C. Ordinal Scale for Clinical Improvement

Patient State	Descriptor	Score
Uninfected	No clinical or virological evidence of infection	0
Ambulatory	No limitation of activities	1
	Limitation of activities	2
Hospitalized Mild disease	Hospitalized, no oxygen therapy	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, ECMO	7
Dead	Death	8

Note: World Health Organization (WHO)-8 point ordinal scale