STATISTICAL ANALYSIS PLAN

Study Protocol Number:	**BGB-3111-219**
Study Protocol Title:	A Phase 2, Randomized, Double Blind, Placebo-Controlled Study of Zanubrutinib Treatment in Patients Hospitalized for COVID-19 Infection and Pulmonary Distress
Date:	Mar 3, 2021
Version:	3.1
	NCT04382586

SIGNATURE PAGE

[redacted]

Date: 3/4/2021

[redacted]

Date: 3/3/2021

Approval

[redacted]

Date: 3/4/2021

[redacted]

Date: 3/3/2021

[redacted]

Date: 3/3/2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Table 1: List of abbreviations and definitions of terms

Abbreviation	Term
AEs	Adverse events
BMI	Body mass index
CI	Confidence interval
CSR	Clinical study report
eCRF	Electronic case report form
ECG	Electrocardiogram
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Toxicity Criteria for Adverse Events
[redacted]	[redacted]
PT	Preferred term
Q1, Q3	First quartile, third quartile
SAEs	Serious adverse events
SAP	Statistical analysis plan
SOC	System organ class
SD	Standard deviation
TEAE	Treatment-emergent adverse event
WHO	World Health Organization

1 INTRODUCTION

This statistical analysis plan (SAP) describes the detailed plan for analysis of data in evaluation of safety and efficacy for study BGB-3111-219. This document is based on protocol amendment 4.0 dated 08Jan2021. The analysis plan for the [redacted] analyses are not included in this SAP.

2 STUDY OVERVIEW

Those patients whose health care providers feel that they qualify to participate in this study, who meet inclusion/exclusion criteria and sign an associated informed consent form will be enrolled into the study. These patients will have been hospitalized for COVID-19 infection and associated respiratory distress. Patients who have been on supplemental $O_2 \leq 96$ hours and have pulmonary distress related to COVID-19 infection will be assigned to Cohort 1 and randomized to receive either zanubrutinib plus best supportive care or placebo plus best supportive care. The dose of zanubrutinib will be 320 mg once daily for a maximum of 28 days. Sixty-three patients will be enrolled in this double blind randomized part (Cohort 1) of the study. [redacted]

The coprimary endpoints for this study are *i*) the respiratory failure-free survival rate at Day 28 in Cohort 1 and *ii*) the time to breathing room air in Cohort 1.

Respiratory failure-free survival at Day 28 is defined as not having died or gone into respiratory failure on or prior to Day 28. Respiratory failure is defined by a clinical diagnosis of respiratory failure and initiation of one of the following therapies:

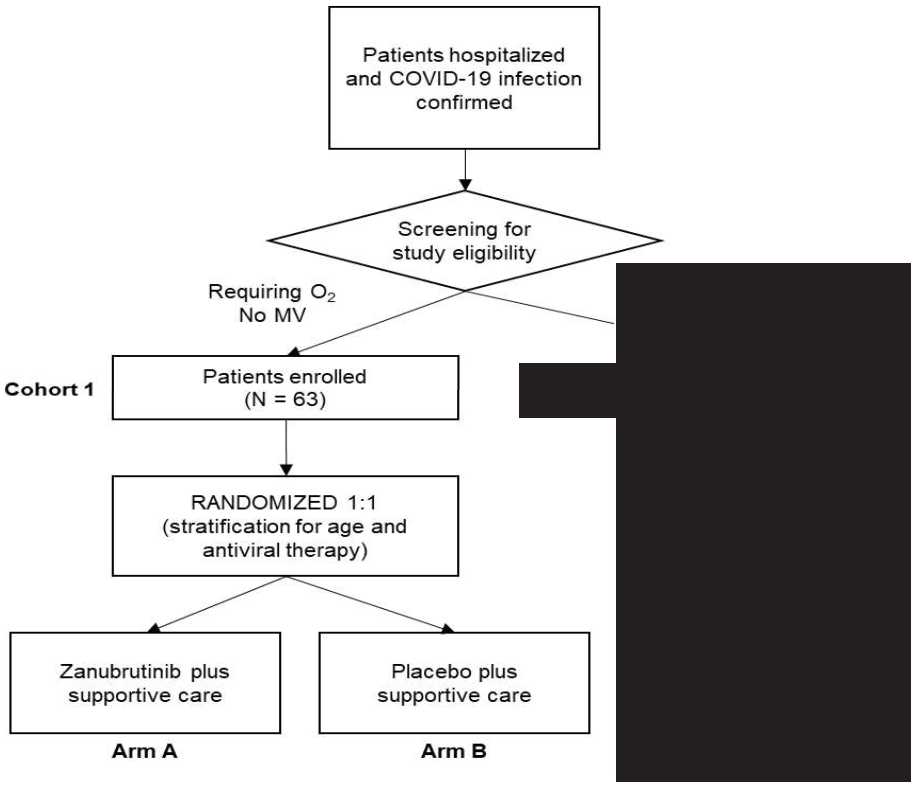
1. Endotracheal intubation and mechanical ventilation,

2. Extracorporeal membrane oxygenation,
3. Clinical diagnosis of respiratory failure with initiation of none of these measures only when clinical decision-making driven is driven solely by resource limitation or changes in goals of therapy (these events will be formally collected in the CRF and flagged, for purposes of sensitivity analyses).

Time to breathing room air is defined as the time from randomization to the earliest time where the patient is stable on room air without any supplemental O_2 therapy. If a patient is on room air for several hours and subsequently receives O_2 therapy, the patient is not considered stable on room air and will not be considered to have returned to breathing room air for purpose of calculating this endpoint. If a patient dies before returning to breathing room air, time to room air will be treated as censored at Day 28 in the analysis. If a patient is lost to follow-up on or before Day 28 without returning to breathing room air, they will be censored at the time they are lost to follow-up. All the patients who do not return to breathing room air on or before Day 28 will be censored at Day 28.

Subjects will be randomized in a 1:1 ratio to zanubrutinib plus best supportive care or placebo plus best supportive care. Randomization will be stratified by Age (> 65 vs <65) and use of anti-viral therapy (Yes vs No). Below is a diagram describing the study.

Figure 1: Study Diagram



3 STUDY OBJECTIVES

3.1 PRIMARY

- To compare the efficacy of zanubrutinib plus supportive care versus placebo plus supportive care as measured by respiratory failure-free survival rate at Day 28 and time to breathing room air (Cohort 1)

3.2 SECONDARY

- To compare the efficacy of zanubrutinib plus supportive care versus placebo plus best supportive care (Cohort 1) as measured by:

- Proportion of patients experiencing respiratory failure or death
- All-cause mortality
- Proportion of patients discharged alive
- Proportion of patients discharged from the ICU alive
- Days spent on supplemental O_2
- Mechanical ventilation-free survival
- Days on mechanical ventilation
- Duration of hospitalization/Time to Discharge
- Change in WHO 8-point scale from baseline
- P_{AO_2}:F_{IO_2} and/or oxygenation index, for patients on mechanical ventilation
- Clinical status as assessed using the WHO 8 point ordinal scale at Days 7, 14, 21, and 28

- To compare the safety of zanubrutinib plus supportive care versus placebo plus best supportive care (Cohort 1) as measured by TEAEs, SAEs, TEAEs by grade, and abnormal laboratory findings.

3.3 EXPLORATORY

- [redacted]
- [redacted]

[redacted]

4 STUDY ENDPOINTS

4.1 PRIMARY ENDPOINT

The coprimary endpoints for this study are *i*) the respiratory failure-free survival rate at Day 28 in Cohort 1 and *ii*) the time to breathing room air in Cohort 1.

Respiratory failure-free survival is defined as the time to the first occurrence of either respiratory failure or death.

Respiratory failure is defined by a clinical diagnosis of respiratory failure and initiation of one of the following therapies:

1. Endotracheal intubation and mechanical ventilation,
2. Extracorporeal membrane oxygenation,
3. Clinical diagnosis of respiratory failure with initiation of none of these measures only when clinical decision-making driven is driven solely by resource limitation or change in goals of therapy (these events will be formally collected in the CRF and flagged, for purposes of sensitivity analyses).

Time to breathing room air is defined as the time from randomization to the earliest time where the patient is stable on room air without any supplemental O_2 therapy. If a patient is on room air for several hours and subsequently receives O_2 therapy, the patient is not considered stable on room air and will not be considered to have returned to breathing room air for purpose of calculating this endpoint. If a patient dies before returning to breathing room air on or before Day 28, time to room air will be treated as censored at Day 28 in the analysis. If a patient is lost to follow-up on or before Day 28 without returning to breathing room air, they will be censored at the time they are lost to follow-up. All the patients who do not return to breathing room air on or before Day 28 will be censored at Day 28.

4.2 SECONDARY ENDPOINTS

Secondary endpoints for Cohort 1 will include:

- Proportion of patients experiencing respiratory failure or death on Study Days 7, 14 and 21
- All-cause mortality at Study Days 7, 14, 21, and 28
- Proportion of patients discharged alive at Study Days 7, 14, 21, and 28
- Proportion of patients discharged from the ICU alive at Study Days 7, 14, 21 and 28
- Days spent on supplemental O_2
- Mechanical ventilation-free survival
- Days on mechanical ventilation.

- Duration of hospitalization/Time to discharge. Death will be assigned the maximum value observed in the trial + 1 and comparison between the two arms will be done using the Wilcoxon rank sum test.
- Change in WHO 8-point scale from baseline. Comparison between the two arms will be done using the Wilcoxon rank sum test. The WHO 8-point scale is defined as in Table 2.
- Proportion of patients whose WHO 8-point scale changed from ≥ 4 at baseline to ≤ 3 at Study Days 7, 14, 21 and 28

Table 2: WHO 8 point ordinal scale for clinical improvement

Patient State	Descriptor	Score
Uninfected	No clinical or virologic evidence of disease	0
Ambulatory	No limitation of activity	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

Abbreviations: RRT, registered respiratory therapist; ECMO, extracorporeal membrane oxygenation.

- P_{AO_2}:F_{IO_2} and/or oxygenation index, for patients on mechanical ventilation.
- Safety and tolerability of zanubrutinib as an adjuvant therapy to standard of care as assessed by TEAEs, SAEs, TEAEs by grade, and abnormal laboratory findings.

Frequency, seriousness and relatedness of TEAEs will be analyzed according to MedDRA. Laboratory abnormalities will be analyzed according to NCI-CTCAE v5.0.

4.3 EXPLORATORY

- [redacted]
- [redacted]
- [redacted]

5 SAMPLE SIZE CONSIDERATION

The study will enroll approximately 67 to 73 patients with 63 patients in Cohort 1 randomized to either zanubrutinib and supportive care (Arm A) or placebo and supportive care (Arm B). [redacted]

Enrollment was anticipated to be complete within 3 to 6 months. The study was expected to be complete within 6 to 9 months after activation.

Below we present the original rational for the size of Cohort 1 [redacted]. At the end of this section we provide a justification for a 50 % increase in the size of Cohort 1.

5.1 SAMPLE SIZE CONSIDERATION FOR COHORT 1

From the SOC arm of the lopinavir-ritonavir study (n = 100) (Cao *et al*, 2020), we have the following statistics regarding respiratory failure and death (Table 3).

Table 3: Respiratory failure and death in the standard care arm of Lopinavir-Ritonavir study

Time point	Respiratory failure n (%)	Death n (%)	Respiratory failure or death n (%)
Baseline	16 (16)		16 (16)
Day 7	25 (25)	7 (7)	32 (32)
Day 14	11 (11)	17 (17)	28 (28)
Day 28		25(25)	NA

Abbreviation: NA, not available.

The maximum observed rate of respiratory failure or death is 32%. In Cohort 1 of the current study under consideration, patients in respiratory failure are not enrolled. However in this study by Cao et al, there were patients in respiratory failure at baseline. Removing patients in the Cao et al study

in respiratory failure at baseline from our estimate of the respiratory failure-free survival rate at Day 28 we estimate the respiratory failure or death rate to be 16% at Day 28 for the population in Cohort 1 of this study or a 84% respiratory failure-free rate.

Myers and others (Myers *et al*, 2020) identified of 377 patients treated for COVID-19 as inpatients at Kaiser Permanente Northern California. The following statistics on the highest level of respiratory support were presented.

Table 4: Highest level of respiratory support in Kaiser Permanente Study

Respiratory support	n (%)
Nasal cannula/face mask	150 (39.8)
High-flow oxygen	12 (3.2)
Noninvasive ventilation	8 (2.1)
Invasive Ventilation	110 (29.2)

Forty-six percent, (110 + 8 + 12) / (110 + 8 + 12 + 150), of patients who received some sort of respiratory support were in what the current protocol defines as respiratory failure. Unfortunately, we cannot decipher how many were in respiratory failure when they were admitted to the hospital. So all we can say from this study by Meyers et al is that the maximum that the respiratory failure rate could be in the study we are considering in this protocol with this information in hand is 46 percent. However, if we assume further that the baseline rate of respiratory failure in the study by Meyers et al is the same as the baseline rate that was observed in the Lopinavir-Ritonavir trial,

Cao et al NEJM March 18 2020, then we would come up with a Respiratory failure rate of 46% - 16% = 30%.

From above, both 70% and 84% were assumed as the respiratory failure-free survival rate for the control arm (Arm B).

Assuming a 10% dropout prior to Day 28, unrelated to efficacy, 38 patients total will provide approximately 81% power to detect an increase in the respiratory failure-free survival rate from 70 to 95% under a z-test for proportions with type 1 error of 0.10 1-sided. It will also provide 77% power for detecting an increase from 80 to 99% and 64% power for detecting an increase from 85 to 99% both with type 1 error of 0.10 one sided.

Table 5: Power for the primary endpoint of respiratory failure-free survival rate at Day 28

Respiratory Failure-Free Survival Rate		**Power (Alpha = 0.10, one-sided, n = 19 per group)**
Control	**Zanubrutinib**	
0.85	0.90	0.208
0.85	0.95	0.405
0.85	0.99	0.642
0.80	0.90	0.341
0.80	0.95	0.561
0.80	0.99	0.767
0.70	0.80	0.286
0.70	0.90	0.622
0.70	0.98	0.897
0.70	0.95	0.807

5.2 SAMPLE SIZE CONSIDERATION FOR [redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

5.3 MODIFICATION TO THE SAMPLE SIZE FOR COHORT 1

During the execution of this study, the RECOVERY study (RECOVERY Collaborative Group et al 2020) reported that the use of dexamethasone reduced the percentage of patients who died by a factor of 0.82 in those patients who received O_2 but not mechanical ventilation. We will assume that the incidence of respiratory failure and death is reduced by a factor of 0.82 since respiratory failure is a precursor event of death. The original analysis plan for Cohort 1 and the data that were used to determine the size of Cohort 1 did not reflect these published data. As a result, the sample size for Cohort 1 will be increased by a factor of 1/0.82**2, which represents an approximately 50% increase in the size of Cohort 1.

$$n = \frac{\left(z_\alpha + z_\beta\right)^2 [p_T(1-p_T) + p_C(1-p_C)]}{(p_T - p_C)^2}$$

Increasing the sample size by 1/0.82**2 accounts for a change in the difference in proportions in the denominator above due to the introduction of dexamethasone. To be conservative, we did not adjust the sample size for the change in the variance component of the sample size formula above.

In addition, an interim/futility analysis will be performed in Cohort 1 at the point when the final analysis for Cohort 1 was previously planned to occur, specifically 28 days after the 42nd randomized patient's first dose date. Table 7 summarizes the planned interim and final analyses for Cohort 1 that will preserve the overall type 1 error at the 0.10, 1-sided level and provide 81% power to detect an increase in respiratory failure-free survival rate at Day 28 from 70% to 91%. The efficacy boundary is determined by the Obrien Fleming spending function. This permits flexibility in the number and timing of interim analyses for this study. The study was initially designed to detect an increase in the respiratory failure-free survival rate from 70% to 95%.

Table 7: Power, sample size, and type 1 error for the interim analysis plan

Parameter	Interim analysis	Final analysis
Number of subjects total	42	63
p-value boundary for efficacy (reject the null)	0.044	0.087
p-value boundary for futility (reject the alternative)	0.15 favoring control	0.087
Cumulative power	0.53	0.807
Cumulative type 1 error (1-sided)	0.044	0.100
Cumulative prob. of stopping for futility (alternative)	0.002	0.193
Cumulative prob. of stopping for futility (null)	0.15	0.90

There is 81% power overall to detect the treatment benefit described above. There is a 53% chance that the study will stop at the interim analysis for efficacy. If the study drug is not effective, the study has a 15% chance to stop for futility at the interim analysis, where futility is defined as a p-value of 0.15 favoring control.

If the respiratory failure-free survival rate at Day 28 is ≤ 0.547 on zanubrutinib compared with 0.700 on control, then the study will be stopped for futility (per an approximation using the z-test).

In the protocol amendment 4, the endpoint time to breathing room air has been made a coprimary endpoint. From the first stage of the Adaptive Covid-19 trial (ACTT-1) which investigated the use of remdesivir, patients on O_2 therapy who received remdesivir took a median of 7 days to recover. Assuming that the treatment effect of Zanubrutinib corresponds to hazard ratio of 0.50 (the event is back to room air), then 38 events will provide 80% power with a type 1 error of 0.10 one sided. A hazard ratio of 0.57 will require 59 events for 80% power with a type 1 error of 0.10 one sided.

6 STATISTICAL METHODS

6.1 ANALYSIS SETS

The Intent-to-Treat (ITT) Analysis Set includes all randomized patients in Cohort 1. The ITT Analysis Set will be the primary population for efficacy analyses in Cohort 1. Subjects will be analyzed according to their randomized treatment assignment.

[redacted]

The Safety Analysis Set includes all patients who received any dose of study drug. Patients will be included in the treatment group corresponding to the actual treatment received. The Safety Analysis Set will be used for all safety analyses.

The Per-Protocol Analysis Set includes patients who received any dose of study medication in Cohort 1 and may exclude those with critical protocol deviations. Criteria for exclusion from the Per-Protocol Analysis Set will be determined and documented before the database lock for the primary analysis.

[redacted]

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

Descriptive statistics include n, mean, standard deviation (SD), median, first quartile (Q1), third quartile (Q3), minimum (min), and maximum (max) for continuous variables and n (%) for categorical variables.

All calculations and analyses will be conducted using SAS version 9.2 or higher or R version 3.6.0 or higher.

6.2.1 Definitions and Computations

Study treatment (study drug): Study drug for this study is zanubrutinib or placebo.

Study Day (Day): Study day will be calculated relative to the date of the first dose of study treatment (Study Day 1). For subjects not dosed, the randomization date will be used instead of the first dose date. For assessments conducted on or after Study Day 1, Study Day will be calculated as (assessment date – Study Day 1 + 1). For assessments conducted before the Study Day 1, Study Day is calculated as (assessment date – Study Day 1). There is no Study Day 0.

In the situation where the event date is partial or missing, the date will appear partial or missing in the listings. Study Day and any corresponding durations will be presented based on the imputations specified in Analysis Details Specification document.

Treatment duration: The treatment duration will be calculated as (date of last dose of study treatment – Study Day 1 + 1).

Baseline: Unless otherwise specified, a baseline value is defined as the last non-missing value collected before Study Day 1. Note that assessments that occur on the first day of dosing but prior to the time of the first dose can qualify to be a baseline value.

6.2.2 Handling of Missing Data

Missing data will not be imputed unless otherwise specified. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior/concomitant medications/procedures.

When summarizing categorical variables, subjects with missing data are generally included in the denominator to calculate percentages unless otherwise specified. When needed, the category of "Missing" is created and the number of subjects with missing data is presented.

When summarizing continuous variables, subjects with missing data are not included in calculations unless otherwise specified.

No imputation of AE grades will be performed. TEAEs with missing CTCAE grade will only be summarized in the all-grades column.

If the assessment of the relationship of an AE to study treatments is missing, then the AE is assumed to be related to the study treatment in the safety analysis summary. No imputation will be done in the AE listings.

6.2.3 Adjustment for Covariates

Covariate adjusted analysis will be performed to adjust for important baseline covariates for the primary and some secondary endpoints. These important baseline covariates could be the stratification factors in the randomization as well as important prognostic factors that are imbalanced between the arm.

6.2.4 Multiplicity Adjustment

The level of the primary hypothesis for this phase 2 study is 1-sided 0.1. There are no planned adjustments for multiplicity.

6.2.5 Data Integrity

Before pre-specified statistical analysis begins, the integrity of the data should be reviewed to assure fit-for-purpose. The data set for analysis should be an accurate and complete representation of the subjects' relevant outcomes from the clinical database. All essential data should be complete and reviewed up to a pre-specified cutoff date. Critical consistency checks and appropriate source data verification should be completed according to the final data extraction plan.

6.3 SUBJECT CHARACTERISTICS

6.3.1 Subject Disposition

The number of patients screened, randomized, treated, prematurely discontinued from study drug (defined as those who discontinued study drug due to any reason except for being discharged and

clinically stable on room air for ≥ 48 hours with an O_2 saturation ≥ 94%) with reasons, completed study treatment, and remained on study treatment will be summarized. In addition the number of patients discontinued from study with reasons and completed study will be counted.

6.3.2 Demographics and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized using the ITT Analysis Set (Cohort 1) [redacted] using descriptive statistics. Continuous variables include age, weight, height, BMI; categorical variables include sex, age group, race, ethnicity, obesity. A listing of demographic and baseline characteristics will be also provided.

6.3.3 Clinical Characteristics at Baseline

A wide range of clinical characteristics at baseline will be summarized with descriptive statistics, primarily counts and proportions. These clinical characteristics include medical history, vital signs, selected laboratory test result, coexisting conditions, symptoms and the WHO 8 point ordinal scale.

6.3.4 Prior and Concomitant Therapy

Prior and concomitant medications will be coded using the WHO Drug Dictionary and the Anatomical Therapeutic Chemical Classification. Prior medications will be defined as medications that stopped before the first dose of study drug. Concomitant medications will be defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 30 days after the patient's last dose.

6.4 EFFICACY ANALYSIS

6.4.1 Primary Efficacy Endpoint Analysis

The primary efficacy analyses for this study involve two comparisons. First a comparison of the respiratory failure-free survival rate at Day 28 between the zanubrutinib plus best supportive care

arm and the placebo plus best supportive care arm and second a comparison of the time to breathing room air for the same two arms in Cohort 1.

Table 8 provides rules for determining respiratory failure-free survival rate at Day 28 as a simple proportion. The rate of respiratory failure-free survival at Days 7, 14 and 21 will be determined similarly. Analyses of respiratory failure-free survival rate is restricted to the time period up to Day 28.

Table 8: Rules for determining respiratory failure-free survival rate at Day 28

Condition	**Rules for determining respiratory failure-free survival rate at Day 28**
Completed Day 28 assessment without any event (death or respiratory failure)	No failure
Death at any time on or before Day 28	Failure
Respiratory failure at any time on or before Day 28	Failure
Discharged from the hospital and lost to follow-up with no record of death nor readmission to the hospital on or before Day 28	No failure
Not discharged from the hospital but lost to follow-up with no record of death or respiratory failure at any time on or before Day 28	If the last measured post-baseline O_2 use measured in liters is less than baseline, treat as no failure. Otherwise, treat as a failure.
Randomized and not treated, discharged or went back to room air before lost to follow-up with no record of death nor readmission to the hospital on or before Day 28	No failure

Table 9 provides censoring rules for the respiratory failure-free survival analysis.

Table 9: Censoring rules for the respiratory failure-free survival analysis

Condition	**Censoring rules for the respiratory failure-free survival analysis**
Completed Day 28 assessment without any event (death or respiratory failure) at any time on or before Day 28	Censored at Day 28
Death at any time on or before Day 28	Failure on death date
Respiratory failure at any time on or before Day 28	Failure on respiratory failure date
Discharged from the hospital and lost to follow-up with no record of death nor readmission to the hospital on or before Day 28	Censored at Day 28
Not discharged from the hospital but lost to follow-up with no record of death or respiratory failure at any time on or before Day 28	Censored on the minimum of Day 28 and the lost to follow-up date
Randomized and not treated, discharged with no record of death nor readmission to the hospital on or before Day 28	Censored on the minimum of Day 28 and the lost to follow-up date

Note in Table 8 that patients who are discharged from the hospital and who are lost to follow-up will be treated as no failure at Day 28. The rational for this is that having been discharged from the hospital, it is very likely that they will be without respiratory failure at Day 28.

Further note that patients who are not discharged but lost to follow-up and their last assessment showed improvement in O_2 use from baseline will be treated as no failure at Day 28. This situation is very unlikely but may occur in 1 or 2 cases.

Assuming that patients who are alive without respiratory failure at their last assessment prior to being lost to follow-up have experienced respiratory failure or death at Day 28 seems hard to justify. Further assuming that patients who are alive without respiratory failure at their last assessment prior to being lost to follow-up are still alive without respiratory failure at Day 28 seems hard to justify as well from the analysis perspective because we will be assuming a good outcome for treatment which is what we are trying to prove. Thus we took an intermediate path and based the imputation of response at Day 28 on the change in O_2 use from baseline. Note that these cases where subjects are lost to follow-up without being discharged from the hospital should be very few in number.

Table 10 provides censoring rules for survival analysis of the time to breathing room air.

Table 10: Censoring rules for survival analysis of the time to breathing room air

Condition	**Censoring rules for survival analysis of the time to breathing room air**
Returned to breathing room air on or before Day 28	Event on the date returned to breathing room air
Completed Day 28 assessment without returning to breathing room air at any time on or before Day 28	Censored at Day 28
Death before returning to breathing room air at any time on or before Day 28	Censored at Day 28
Discharged from the hospital and lost to follow-up with no record of returning to breathing room air on or before Day 28	Censored at lost to follow-up date
Not discharged from the hospital but lost to follow-up with no record of returning to breathing room air on or before Day 28	Censored at lost to follow-up date

Randomized and not treated, discharged with no record of returning to breathing room air on or before Day 28	Censored at hospital discharge date

The null and alternative hypotheses for formally testing the superiority of zanubrutinib over placebo for the endpoint respiratory failure-free survival rate at Day 28 are as follows:

H_0: $p_{zanu} = p_{placebo}$

H_a: $p_{zanu} > p_{placebo}$

where p denotes the rate of respiratory failure free survival at Day 28, that is the proportion of patients who have not experienced respiratory failure or death at any time on or before Day 28. The primary inference regarding this hypothesis will be based on unstratified Fisher's exact test with a type 1 error rate of 0.10 1-sided. The associated odds ratio and its 2-sided 95% CI will also be presented. If the 1-sided p-value is less than 0.10, the addition of zanubrutinib to best supportive care will be considered to increase the respiratory failure-free survival rate at Day 28. Because testing will be carried out at the 0.10 1-sided level of significance instead of the traditional 0.05 two-sided level, a positive result will not be considered definitive evidence of zanubrutinib's effectiveness.

The null and alternative hypotheses for formally testing the superiority of zanubrutinib over placebo for the endpoint time to breathing room air are as follows:

H_0: $\lambda = 1$

H_a: $\lambda > 1$

where $\lambda = \lambda_{zanu}/\lambda_{placebo}$ represents the hazard ratio of the event return to breathing room air between zanubrutinib (hazard λ_{zanu}) and placebo (hazard $\lambda_{placebo}$) arms. The impact of adding zanubrutinib to best supportive care on the time to breathing room air will be formally

tested using the hazard ratio from a Cox model with treatment as an explanatory variable. P-value will be calculated from the 2-sided log-rank test. For exploratory analysis purpose, the stratification variables of age category (< 65 or ≥ 65) and use of antivirals (yes or no) may be included as additional explanatory variables in the Cox model.

Two analyses (interim and final) will be performed for the coprimary endpoints in Cohort 1. The interim analysis will summarize Day 28 follow-up of the first 42 randomized patients and will be performed no sooner than 28 days after the 42nd randomized patient's first dose date. The final analysis will be a complete summary of the efficacy and safety data for all patients randomized to the study and will take place no sooner than 84 days (28 days of treatment and 56 days of safety follow-up) after the last randomized patient's first dose date. The type 1 error increase due to these 2 analyses will be controlled by O'Brien Fleming alpha spending functions. The boundary for efficacy stop will be based on two p-values, one from the Fisher's exact test for the respiratory failure-free survival rate at Day 28 and the other from Cox model for time to breathing room air.

Potential intercurrent events that can affect the estimates of the rates of respiratory failure-free survival at Day 28 include treatment discontinuation due to an adverse event and lost to follow-up after hospital discharge. Regarding treatment discontinuation there will be no adjustments for this intercurrent event since the intention is to estimate the effect of zanubrutinib in this clinical setting and that includes outcomes for patients who experience an adverse event and as a result have to stop taking zanubrutinib. Thus the outcome for these patients will depend on further follow-up for respiratory failure or death.

Regarding subjects who are lost to follow-up following hospital discharge, we will treat these subjects as alive and without respiratory failure up to Day 28 unless they die or are re-hospitalized with respiratory failure. However subjects who are lost to follow-up without hospital discharge and without respiratory failure or death will be censored at the point where they are lost to follow-up for purposes of survival analyses. For purposes of determining the respiratory failure-free

survival rate as a simple proportion these patients will be handled as indicated in Table 8 depending on the change in their O_2 support from baseline. As supportive analyses, we intend to utilize survival analysis techniques that incorporate methods for censoring to adjust estimates of respiratory failure-free survival at Days, 7, 14, 21 and 28 for subjects who are lost to follow-up if there are any. Testing based on survival analysis techniques would then rely on Kaplan Meier estimates and Greenwood's formula for the variance of these estimates.

It should be noted that death is an expected intercurrent event for time to breathing room air since it does not numerically incorporate death. It is our intention to use time to breathing room air to compare zanubrutinib's effectiveness with placebo. Therefore we must account for deaths somehow. As noted for the formal comparison of the two arms of the trial with this endpoint, a Cox model is used where patients who have died without returning to breathing room air are censored at Day 28.

The impact of treatment on respiratory failure-free survival will also be assessed with an exact conditional logistic regression model. Because we only expect to have 10 of 63 subjects with respiratory failure, these analyses will be exploratory in nature.

The impact of treatment on respiratory failure-free survival will be also compared between the two arms using a Cox proportional hazards regression approach, treating the occurrence of respiratory failure or death as an event. Time to an event will be measured starting from the randomization date up to Day 28. For exploratory analysis purpose, the stratification variables of age category (< 65 or ≥ 65) and use of antivirals (yes or no) may be included as explanatory variables. These analyses will be exploratory in nature since we only expect to have around 10 of 63 subjects with respiratory failure or death.

Demographic and baseline disease characteristics will be assessed by randomization group to identify any prognostic factors in which there was an imbalance between the treatment groups.

Should an imbalance exist in spite of randomization, the variable in which an imbalance is noted, may be added as an explanatory variable to adjust for the imbalance in the efficacy interpretation.

6.4.2 Secondary Efficacy Endpoints Analysis

The types of secondary efficacy endpoints fall into three general categories. Comparisons of rates, comparisons of continuous measurements and comparisons of times to an event.

6.4.2.1 Comparisons of rates

The following secondary endpoints fall into the general category of comparisons of rates.

- Proportion of patients experiencing respiratory failure or death on Study Days 7, 14 and 21
- All-cause mortality at Study Days 7, 14, 21, and 28
- Proportion of patients discharged alive at Days 7, 14, 21, and 28
- Proportion of patients discharged from the ICU alive at Days 7, 14, 21 and 28

For these endpoints, the number and percent of subjects with an event will be presented. An unstratified 1-sided Fisher's exact test will be used to compare the two groups on these endpoints and the associated odds ratio and 95% CI will be provided. Note that for the proportion of patients discharged alive from the ICU, the denominator will be the cumulative number of patients admitted to the ICU on or before Days 7, 14, 21, and 28 respectively. The denominator will be presented along with the cumulative number of patients discharged alive from the ICU. Because this endpoint is based on a subset of the randomized population any differences observed between the arms should be interpreted cautiously.

Possible intercurrent events that can affect the estimate of these endpoints are discontinuation of study drug due to adverse events, and lost to follow-up after being discharged from the hospital for the endpoints respiratory failure or death, and all cause mortality. Regarding discontinuation of study drug due to adverse events, the intention is to estimate the effect of zanubrutinib in this

clinical setting and that includes outcomes for patients who experience adverse events and can no longer take zanubrutinib. Consequently, this intercurrent event will not lead to adjustments in the estimates of the rates of these endpoints. As far as lost to follow-up following hospital discharge is concerned, the intention is to estimate the respiratory failure-free survival and all cause mortality for all subjects. Subjects who are discharged alive and are lost to follow-up will be treated as alive in these analyses at all later timepoint. Similarly for subjects who are discharged without respiratory failure up to that point and lost to follow-up without observing the endpoint following hospital discharge, survival analysis techniques may be used to adjust the estimate of these endpoints. Such subjects will be censored at the point where they are lost to follow-up. In the situation where subjects are discharged from the hospital, they will be assumed alive and respiratory failure-free at all later time points unless data is collected to the contrary. Comparison of the Kaplan Meier estimates in the two arms at Days 7, 14, 21 and 28 will be made with the assistance of Greenwood's formula for the variance of these estimates.

6.4.2.2 Comparisons of continuous measurements

The following endpoint analyses are based on comparisons of continuous measurements.

- Days on mechanical ventilation
- Duration of hospitalization/Time to discharge

It should be noted that death is an expected intercurrent event for these endpoints since they do not numerically incorporate death. It is our intention to use these endpoints to compare zanubrutinib's effectiveness with placebo. Therefore we must account for deaths somehow. For the formal comparison of the two arms of the trial with these endpoints, a Wilcoxon rank sum test will be used where patients who have died without an event are assigned a value equal to 1 plus the maximum value observed in Cohort 1. That is deaths will be assigned to the highest/lowest rank which represents the worst patient outcome. The actual observed values of these endpoints will be further summarized with descriptive statistics by treatment arm. In addition to the rank-based

Wilcoxon test, we will apply a test of the hypothesis that the medians are the same developed by Mood (Mood *et al*, 1954). This test involves creating a 2x2 table of treatment by subjects who are above *vs* equal to or below the overall median and testing with a chi-square test. Note that we anticipate the death rate in Cohort 1 to be less than 50%. If this is the case then assigning patients who died the worst rank will allow us to estimate a true median value for these endpoints and Moods test on the medians will allow us to assess if any differences we observe in the medians are statistically significant.

Other intercurrent events that could affect these endpoints are stopping treatment due to an adverse event. Our intention is to describe the impact of zanubrutinib in this clinical setting and stopping treatment due to an adverse event is part of the clinical setting. So no adjustment for this will be required.

6.4.2.3 Comparisons of the time to an event

The following endpoints are based on comparisons of time to an event.

- Respiratory failure-free survival
- Mechanical ventilation-free survival

Both respiratory failure-free survival and mechanical ventilation-free survival are endpoints that will be evaluated using traditional survival analysis techniques. Subjects will be censored as in Table 9 if they have not had an event prior to the data cutoff or if they are lost to follow-up. The number of patients with respiratory failure/mechanical ventilation, and death as well as the total follow-up time will be summarized. Simple estimates of the hazard (total number of events/total follow-up time) and the hazard ratio (95% CI) from a Cox model with treatment as the sole explanatory variable will be provided. P-value will be calculated from the 2-sided log-rank test.

6.4.2.4 Other Secondary Efficacy Endpoints

6.4.2.4.1 WHO 8 ordinal scale for clinical improvement

The actual values and the change from baseline in WHO 8-point scale will be summarized by descriptive statistics. Comparison of the change from baseline between the two arms will be done using the Wilcoxon rank sum test. The WHO 8-point scale is defined as in Table 6 in the protocol. In addition shift tables will be created for comparing baseline with Days 7, 14, 21 and 28.

Proportion of patients whose WHO 8-point scale changed from ≥ 4 at baseline to ≤ 3 at Study Days 7, 14, 21 and 28 will be compared between two arms using unstratified 1-sided Fisher's exact test and the associated odds ratio and 95% CI will be provided.

6.4.2.4.2 PAO_2:FIO_2/Oxygenation Index

For patients in Cohort 1 who started mechanical ventilation, PAO_2:FIO_2/Oxygenation index appropriately standardized will be calculated and listed.

6.4.2.4.3 [redacted]

6.4.3 Exploratory Efficacy Analysis

[redacted]

6.4.4 Sensitivity Analysis

For the coprimary endpoint of the respiratory failure-free survival rate at Day 28, a logistic regression analysis will be undertaken as a complement to the unstratified primary analysis. The explanatory variables will be those incorporated in the randomization, age (>= 65 vs < 65) and use of antivitral therapy (yes vs no). In addition, baseline factors that are associated with respiratory

failure-free survival and are imbalanced between the treatment arms may be included in the logistic regression analyses to see how they impact the estimate of the treatment effect.

Similar sensitivity analyses may be undertaken for secondary endpoints that involve the comparison of rates at specific time points.

For coprimary efficacy endpoints, a Per-Protocol Analysis may be undertaken to assess the robustness of the result obtained from the ITT analysis.

6.4.5 [redacted]

[redacted]

6.5 SAFETY ANALYSES

Safety will be assessed by monitoring and recording of all AEs graded by NCI-CTCAE v5.0. Laboratory values (CBC, serum chemistry and coagulation), vital signs, physical exams and ECG findings will also be used in assessing safety. Descriptive statistics will be used to analyze all safety data by treatment group.

6.5.1 Extent of Exposure

Extent of exposure to study drug will be summarized descriptively as the duration of exposure (days), cumulative total dose received per patient (mg), dose intensity (mg/day) and relative dose intensity (%).

The number of patients with dose reductions will be summarized with simple proportions. The number of dose reductions per patient will be summarized by numerical category (e.g. 1, 2, 3, or ≥ 4). Doses missed will be summarized by counts as well as by descriptive statistics.

Patient data listings will be provided for all dosing records and for calculated summary statistics.

6.5.2 Adverse Events

The AE verbatim descriptions (as recorded by the investigator on the eCRF) will be classified into standardized medical terminology using Medical Dictionary for Regulatory Activities (MedDRA®). Adverse events will be coded to MedDRA (Version 23.0) lower level term closest to the verbatim term. The linked MedDRA preferred term and primary system organ class will also be captured in the database.

A treatment-emergent AE (TEAE) is defined as an AE that had an onset date on or after the first dose date (Day 1) of study drug (zanubrutinib or placebo) up to 56 days following the last dose date.

- The incidence of TEAEs will be reported as the number and percentage of patients with TEAEs by System Organ Class (SOC) and Preferred Term (PT). A patient will be counted only once by the highest severity grade according to NCI-CTCAE v5.0 within a system organ class and preferred term, even if the patient experienced more than 1 TEAEs. An overall summary of TEAEs will be presented by treatment groups for the number of patients with (1) any TEAE, (2) any Grade 3 or higher TEAE, (3) any serious TEAE, (4) any treatment-related TEAE, (5) any TEAE leading to dose modification (interruption and reduction), (6) any TEAE leading to study treatment discontinuation, (7) any TEAE leading to death. A general summary of TEAEs by SOC and PT will be provided with the number and percent of patients who experienced the following types of events by treatment groups:
- Any TEAE
- Any Grade 3 or higher TEAE
- Any serious TEAE
- Any treatment-related TEAE
- Any TEAE leading to dose modification

- Any TEAE leading to study treatment discontinuation
- Any TEAE leading to death
- Any TEAE of special interest

Data listing for all adverse events will be provided.

6.5.2.1 AE's of Special Interest

AE's of special interest, i.e. adverse events that are generally related to the drugs mechanism of action, will be summarized. The definitions of AE's of special interest are provided in Appendix A.

6.5.3 Laboratory Analyses

CBC and serum chemistry values will be evaluated for each laboratory parameter by treatment group. Abnormal laboratory values will be flagged and identified as those outside (above or below) the normal range. Reference (normal) ranges for laboratory parameters will be included in the CSR. Descriptive summary statistics for laboratory parameters and their changes from baseline will be calculated. Laboratory values will be summarized by visit and by worst post-baseline visit.

Laboratory parameters that are graded in NCI-CTCAE v5.0 will be summarized by CTCAE grade. In the summary of laboratory parameters by CTCAE grade, parameters with CTCAE grading in both high and low directions (e.g. calcium, glucose, magnesium, potassium, sodium) will be summarized separately.

Laboratory measurements that are collected without units will not be included in the summary tables described above. Instead these lab assessments will be identified in a listing of labs with missing units.

6.5.4 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure, heart rate, body temperature, and weight) and changes from baseline will be presented by visit and treatment group for all visits.

6.5.5 Electrocardiogram

Descriptive statistics for ECG assessments and changes from baseline will be presented by visit and treatment group for all visits.

7 INTERIM ANALYSIS

7.1 INTERIM ANALYSIS FOR COHORT 1

When 42 subjects are randomized, an interim analysis will take place 28 days after the 42nd randomized patient's first dose date. For the interim analysis in Cohort 1, only the first 42 randomized patients will be included.

Type 1 error will be formally controlled using an O'Brien Fleming spending functions where the maximum information is 63 subjects. A futility analysis will be undertaken at that time as well. If the p-value for a test of respiratory failure-free survival favors control with a p-value of 0.15, then the futility boundary will have been crossed. These boundaries for the interim analysis are guides to help the SMC in their decision making.

7.2 INTERIM ANALYSIS FOR [redacted]

[redacted]

7.3 FINAL ANALYSIS

The final analysis will take place 84 days after the last randomized patient's first dose date. The analysis of the coprimary endpoint will be based on Day 28 follow-up for each randomized patient in Cohort 1.

The study will be frequently monitored by a Safety Monitoring Committee on a every 1 to 2 week basis for the purpose of monitoring safety.

8 DEVIATIONS FROM THE PROTOCOL

The following table identifies items where this SAP differs from the protocol.

Table 11: Table of difference between this SAP and the protocol

Protocol	SAP	Rationale
Duration of Hospitalization and Time to Discharge are two separate secondary objectives	Duration of Hospitalization and Time to Discharge are combined into 1 secondary objective	These are two terms describing the same quantity
Safety as assessed by related AE's is a secondary objective	[redacted]	In cohort 1 there is a randomized control so the investigator assessment of relatedness is not as informative as in a single arm trial.
Secondary Objective: Clinical Stats assessed using an ordinal	Secondary Objective: Replaced with Clinical Status as assessed	More specific

scale at a pre-specified timepoint	using the WHO 8 point ordinal scale at Days 7, 14, and 28	
Objective: Time to objective measure of recovery	Deleted	Redundant. Time to return to breathing room air was intended to replace "Time to objective measure of recovery".
The protocol describes the use of a time dependent covariate for dosing in a model of Respiratory failure-free survival as an important supportive analysis in support of the primary endpoint	The SAP describes the use of this time dependent covariate for dosing as an exploratory analysis. A cox model with just the stratification factors is the main supportive analysis for the primary endpoint	It is not appropriate to adjust a main analysis of treatment effect with information post baseline. The post baseline information can inappropriately alter the estimate of the treatment effect.
Includes related AEs as a secondary safety endpoint in cohort 1	[redacted]	Cohort 1 has a randomized control arm.
Endpoint: Clinical Stats assessed using an ordinal scale at a pre-specified timepoint	Endpoint: Replaced with Clinical Status as assessed using the WHO 8 point ordinal scale at Days 7, 14, and 28	More specific
Endpoint: Time to objective measure of recovery	Deleted	Redundant. Time to return to breathing room air was intended to replace "Time to objective measure of recovery".

Covariate adjusted analysis will be performed to adjust for important baseline covariates for the primary and some secondary endpoints.	Covariate adjusted analysis will be performed to adjust for important baseline covariates for the primary and some secondary endpoints. These important baseline covariates could be the stratification factors in the randomization as well as important prognostic factors that are imbalanced between the arms.	Added text to describe what could be used as independent variables in the model.
Nothing on intercurrent events for the efficacy endpoints.	Added test for intercurrent events and describe the use of Moods test along with the difference in medians as another test to be performed in addition to the Wilcoxon rank sum test.	To respond to FDA comments on the protocol
The number (percentage) of patients with TEAEs will also be summarized by relationship to study drug	This will be done just for arm 1	
ECOG performance status will be summarized at each visit, etc.	Deleted	we are not collecting ECOG performance status

Text indicated death assumes the maximum value observed in the trial plus 1	Changed to death assumes the minimum value observed in the trial – 1 for PAO_2/FIO_2	Low values for PAO_2:FIO_2 represent a worse outcome.
One of the secondary endpoint is "Median reduction in days spent on supplemental oxygen"	Changed to "Days spent on supplemental oxygen"	The original intention was to present the nominal days, not the median reduction.

[redacted]

Protocol

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

SAP

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

9 REFERENCES

Cao B, Wang Y, Wen D, et al. (2020) A trial of Lopinavir-Ritonavir in adults hospitalized with severe Covid-19. *N. Engl. J. Med*., 382:1787-1799.

Myers L, Paorodi S, Escobar G et al. (2020) Characteristics of hospitalized adults with COVID-19 in an integrated health care system in California. *JAMA*, 323 (21): 2195-2198.

Richardson S, Hirsch J, Narasimhan M, et al. (2020) Presenting characteristics, comorbidities and outcomes among 5700 patients hospitalized with Covid-19 in the New York city area. *JAMA,* 323 (20): 2052-2059.

Mood, A. M. (1954) On the asymptotic efficiency of certain non-parametric two sample tests. *Ann. Math. Statist.*, 25 (3): 514-522.

Simon R. (1989) Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*, 10 (1):1-10.

APPENDIX A: ADVERSE EVENTS OF SPECIAL INTEREST

Adverse Event of Clinical Interest Category	Search Criteria
Hemorrhage	Haemorrhage terms (excl laboratory terms) (SMQ) Narrow
Major hemorrhage	Major hemorrhage = All Haemorrhage PT if AE SOC is 'Nervous system disorders' or serious or grade 3 and above Haemorrhage PT if AESOC is not 'Nervous system disorders'
Atrial fibrillation and flutter	Atrial fibrillation PT, Atrial flutter PT
Hypertension	Hypertension (SMQ) Narrow
Second primary malignancies	Malignant Tumours (SMQ) Narrow
Tumor lysis syndrome	Tumour lysis syndrome (SMQ) Narrow
Infections Opportunistic infections	Infections: Infections and Infestations SOC Subcategory - Opportunistic infections: Opportunistic infections (CMQ)
Neutropenia	Neutropenia PT, Neutrophil count decreased PT, Febrile neutropenia PT, Agranulocytosis PT, Neutropenic infection PT, Neutropenic sepsis PT
Thrombocytopenia	Thrombocytopenia PT, Platelet count decreased PT
Anemia	Anaemia PT, Haemoglobin decreased PT