REGENERON
Biostatistics and Data Management

STATISTICAL ANALYSIS PLAN FOR PHASE 2 PORTION OF STUDY VERSION: FINAL

AN ADAPTIVE PHASE 2/3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY ASSESSING EFFICACY AND SAFETY OF SARILUMAB FOR HOSPITALIZED PATIENTS WITH COVID-19

Compound:	REGN88 (sarilumab; Kevzara®)
Protocol Number:	6R88-COV-2040
Clinical Phase:	Phase 2/3
Sponsor:	Regeneron Pharmaceuticals, Inc.
Study Biostatisticians:	[redacted]
	[redacted]
Clinical Trial Manager:	[redacted]
Study Medical Directors:	[redacted]
	[redacted]
Version/Date:	Original (Version 1.0) / 13 APR 2020

The approval signatures below indicate that these individuals have reviewed the Statistical Analysis Plan (SAP) and agreed on the planned analysis defined in this document for reporting.

See appended electronic signature page

Study Biostatisticians [redacted]

See appended electronic signature page

Study Pharmacokinetics [redacted]

See appended electronic signature page

Study Medical Directors [redacted]

See appended electronic signature page

Head of BDM [redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Definition
AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
ANC	Absolute Neutrophil Count
ANCOVA	Analysis of covariance
ARDS	Acute respiratory distress syndrome
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
COVID-19	Coronavirus Disease 2019
CPK	Creatine phosphokinase
CRF	Case report form (electronic or paper)
CRP	C-reactive protein
ECG	Electrocardiogram
ECMO	Extracorporeal membrane oxygenation
FiO_2	Fraction inspired oxygen
ICH	International Council for Harmonisation
ICU	Intensive care unit
IDMC	Independent Data Monitoring Committee
IL-6	Interleukin 6
ITT	Intention-to-treat
IWRS	Interactive Web Response System
IV	Intravenous
mITT	Modified intention-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
MERS-CoV	Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV)
NCI-CTCAE	National Cancer Institute-Common Terminology Criteria for Adverse Events
NEWS2	National Early Warning Score2
OP	Oropharyngeal

Abbreviation	Definition
NP	Nasopharyngeal
PCSV	Potentially Clinically Significant Value
PD	Pharmacodynamic(s)
PK	Pharmacokinetic(s)
PPS	Per Protocol population set
PT	Preferred Term
RA	Rheumatoid Arthritis
RBC	Red Blood Cell
RNA	Ribonucleic Acid
Regeneron	Regeneron Pharmaceuticals, Inc.
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SAS	Statistical Analysis System
SC	Subcutaneous
SOC	System organ class
SpO_2	Peripheral capillary oxygen saturation
SUSAR	Suspected unexpected serious adverse reaction
TEAE	Treatment-emergent adverse event
ULN	Upper Limit Normal
US	United States (of America)
VFD	Ventilator-free days
WBC	White blood cell
WHO	World Health Organization
WHODD	World Health Organization Drug Dictionary

1. OVERVIEW

The purpose of the statistical analysis plan (SAP) is to ensure the credibility of the study results by pre-specifying the statistical approaches for the analysis of study data prior to the Phase 2 database lock of this Phase 2/3 adaptive study. This Phase 2 SAP is intended to be a comprehensive and detailed description of the strategy and statistical methods to be used in the analysis of data for Phase 2 portion of the 6R88-COV-2040 study based on Protocol Amendment 4 (dated 07-APR-2020). A separate Phase 3 SAP will be written for the Phase 3 portion of the Phase 2/3 study.

This plan may be revised during the study to accommodate protocol amendments and/or to make changes to adapt to unexpected issues in study execution and/or data that affect planned analyses. These revisions will be based on blinded review of the study and data, and a final plan will be issued prior to the Phase 2 data lock and before code breaking.

1.1. Background/Rationale

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel enveloped RNA betacoronavirus that emerged in December 2019 in Wuhan, China. The term COVID-19 is the disease caused by SARS-CoV-2 with symptoms that manifest a median of 5 days and up to 14 days after infection. Patients with COVID-19 infection are at risk for the development of pneumonia and acute respiratory distress syndrome (ARDS), a syndrome of severe impairment of gas exchange, which clinically presents with severe hypoxemia. The most frequent clinical presentation of severe COVID-19 is pneumonia with symptoms including fever, cough, and dyspnea (shortness of breath).

COVID-19 infection can be associated with a degree of pulmonary cytokine release, with IL-6 being a major contributor to the development of fever and hypoxemia. In the current study, sarilumab (Kevzara®), an anti-IL-6R monoclonal antibody (mAb), 200 mg intravenous (IV) and 400 mg IV doses will be evaluated for efficacy and safety in treatment of COVID-19. Kevzara® (sarilumab) is currently approved for treatment in patients with rheumatoid arthritis (RA) at 200 mg Q2W (subcutaneous [SC]) [with down dosing to 150 mg Q2W (SC) for certain laboratory changes]. Sarilumab is highly similar to tocilizumab and there is reported evidence in a recent study in 21 patients that IV treatment with 400 mg of tocilizumab (Actemra®), another anti-IL-6R mAb, provided a clinically meaningful improvement in clinical symptoms that are thought to be mediated by cytokine release in patients with severe or critical COVID-19 infection (Xu, 2020 [7]).

Currently, however, there are no specific COVID-19 treatments. SARS-CoV-2 that causes COVID-19 results in a similar acute lower respiratory disease caused by SARS-CoV, identified in 2002, and by the Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV). (Hereafter, SARS-CoV2 and COVID-19 are interchangeably used.) Many therapeutic agents have been used to treat patients with SARS-CoV and MERS-CoV, including corticosteroids, type 1 IFN agents, convalescent plasma, ribavirin, lopinavir/ritonavir, proteases, and agents targeting viral entry proteins. However, none have been proven to be efficacious in clinical trials, justifying the use of placebo as a comparator for this study.

This study is an adaptive Phase 2/3, randomized, double-blind, placebo-controlled trial to evaluate the efficacy and safety of sarilumab compared with placebo in hospitalized adults with severe or

critical COVID-19 (as broadly defined below). The study will be conducted in the United States (US) in up to 100 sites.

Patients were to be randomized in a 2:2:1 allocation ratio to sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo. Randomization was to be stratified by:

- Severity of illness at enrollment
  - Severe disease
  - Critical disease
  - Multi-system organ dysfunction
    - Immunocompromised†
- Systemic corticosteroids (Yes/No)

† Immunocompromised patients were permitted to be enrolled in this study after Protocol Amendment #3 (28-MAR-2020) at the request of the FDA. For Phase 2 analysis, these patients will be reported as a subset of the Multi-system organ dysfunction strata.

Due to the novel nature of COVID-19, efficacy endpoints are not well established, and the standard of care is expected to evolve over time. The adaptive design allows for the assessment of efficacy endpoints in Phase 2 which will then be confirmed in the Phase 3 portion of the study, as well as evaluate the benefit-risk of 2 doses of sarilumab (200 mg IV or 400 mg IV). This study is therefore intended to allow for multiple adaptations, including dropping of a sarilumab dose arm, modification of the primary endpoint for Phase 3, and sample size re-estimation for Phase 3.

1.2. Study Objectives

1.2.1. Primary Objectives

Phase 2:

The primary objective of the study is to evaluate the clinical efficacy of sarilumab relative to the control arm in adult patients hospitalized with COVID-19 regardless of severity strata.

Phase 3:

The primary objective of the study is to evaluate the clinical efficacy of sarilumab relative to the control arm in adult patients hospitalized with severe and critical COVID-19.

1.2.2. Secondary Objectives

Phase 2 and Phase 3:

Efficacy

The secondary efficacy objectives of the study are to:

1. Evaluate the clinical efficacy of sarilumab compared to the control arm in all disease severity levels and by clinical severity

2. Evaluate the clinical efficacy of sarilumab compared to the control arm by baseline IL-6 level
3. Evaluate changes in the National Early Warning Score 2 (NEWS2)
4. Evaluate the duration of predefined symptoms and signs (if applicable)
5. Evaluate the duration of supplemental oxygen dependency (if applicable)
6. Evaluate the incidence of new mechanical ventilation use during the study
7. Evaluate the duration of new mechanical ventilation use during the study
8. Evaluate need for admission into intensive care unit (ICU)
9. Evaluate duration of hospitalization (days)
10. Evaluate the 28-day mortality rate

Phase 2 and Phase 3:

Safety

The secondary safety objectives of the study are to evaluate the safety of sarilumab through hospitalization (up to Day 29 if patient is still hospitalized) compared to the control arm as assessed by incidence of:

- Serious adverse events (SAEs)
- Grade 4 neutropenia (ANC < 500/mm^3)
- Grade 4 neutropenia (ANC < 500/mm^3) with concurrent severe or life-threatening bacterial, invasive fungal, or opportunistic infection
- Grade ≥2 infusion-related reactions
- Grade ≥2 hypersensitivity reactions
- Increase in alanine transaminase (ALT) or aspartate aminotransferase (AST) ≥3X upper limit of normal (ULN) (for patients with normal baseline) or >3X ULN AND at least 2-fold increase from baseline value (for patients with abnormal baseline)
- Invasive bacterial or fungal infections of clinical significance with confirmed diagnosis based on the investigator's assessment with appropriate diagnostic workups and consultations

1.2.3. Modifications from the Statistical Section in the Final Protocol

The initial protocol and analysis plan was based on a published case series from China suggesting IL-6R blockade in patients with COVID led to rapid reversal of respiratory compromise. Subsequent data suggest that these reference data may have been highly selective. Therefore, in the absence of reliable antecedent data, it was decided that the Phase 2 portion of the study should focus on describing the differences between treatment with sarilumab on top of standard of care versus standard of care alone. Consistent with this decision, the modification in the SAP compared to the protocol is that in the Phase 2 analysis the key secondary endpoint of time-to-improvement

(≥2 points) in clinical status assessment using the ordinal scale will be analyzed descriptively rather than as formal hypothesis testing (see Section 5.7.2 and Section 5.7.4). The approach to control multiplicity has changed, but overall Type 1 error will still be controlled at 0.05.

1.2.4. Revision History for SAP Amendments

Not applicable.

2. INVESTIGATION PLAN

2.1. Study Design and Randomization

This study is an adaptive Phase 2/3, randomized, double-blind, placebo-controlled trial to evaluate the efficacy and safety of sarilumab in hospitalized adults with severe or critical COVID-19. The study will be conducted in the United States (US) in up to 100 sites. All patients must have evidence of SARS-CoV-2 infection pneumonia by radiograph or clinical findings, and respiratory compromise as demonstrated by a requirement for oxygen supplementation above their background requirements.

Patients will be randomized in a 2:2:1 allocation ratio to sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo, in a stratified manner. Randomization will be performed according to a central randomization scheme using an interactive web response system (IWRS). While serum IL-6 and cytokines will be measured at a central laboratory predose on Day 1 and periodically after treatment, entering patients will not be stratified based on serum IL-6 levels.

Randomization will be stratified by:

- Severity of illness at enrollment
  - Severe disease
  - Critical disease
  - Multi-system organ dysfunction
    - Immunocompromised†
- Systemic corticosteroids (Yes/No)

† Immunocompromised patients were permitted to be enrolled in this study after Protocol Amendment #3 (28-MAR-2020) at the request of the FDA. For Phase 2 analysis, these patients will be reported as a subset of the Multi-system organ dysfunction strata.

The severity categories are:

1. **Severe disease**
   - Requires supplemental oxygen administration by nasal cannula, simple face mask, or other similar oxygen delivery device
2. **Critical disease:**
   - Requires supplemental oxygen requiring delivered by non-rebreather mask or high-flow nasal cannula, OR
   - Use of invasive or non-invasive ventilation, OR
   - Requiring treatment in an intensive care unit.
3. **Multi-system organ dysfunction:**
   - Multi-system organ dysfunction: use of vasopressors, extracorporeal life support, or renal replacement therapy

4. **Immunocompromised**
    - Immunocompromised patients (or on immunosuppressant treatments)

The total sample size for Phase 2 will be approximately 460 to include patients with all baseline IL-6 levels.

For the Phase 3 portion of this study, the primary endpoint will be determined after data from the Phase 2 portion is analyzed by the unblinded study team.

Sample size may be re-estimated for Phase 3 based on results from Phase 2 data.

Figure 1: Study Flow Diagram for Phase 2 and Phase 3

[a] The EOS will be on day 60 or day of death, whichever comes first.

[b] If the patient has been discharged from the hospital before day 29, the study site staff will contact the patient for a follow-up phone call.

[c] If the patient has been discharged from the hospital before day 60, the study site staff will contact the patient for a follow-up phone call.

EOS: End of study

2.2. Sample Size and Power Considerations for Phase 2

For the Phase 2 portion of the study, approximately 200 patients (80 in each of the two sarilumab dose groups and 40 on placebo) are required in all disease severity strata with high baseline IL-6 levels to test for superiority of sarilumab versus placebo with respect to percent change from baseline CRP levels at Day 4. This sample size will provide 90% power using a two-sample t-test to detect an effect size (ie, difference/SD) of 0.633 at the 0.05 (2-sided) significance level. The effect size of 0.633 assumes that the mean treatment difference in percent change from baseline using natural logarithm of CRP levels (sarilumab minus placebo) would be -6.6% with a standard deviation of 10.43% (percent change from baseline in sarilumab is estimated to be mean = 45.98% and SD=10.43% based on tocilizumab data from China study in COVID-19; See Figure 2 in (Xu, 2020 [7])). If the effect in placebo group is assumed to be negligible (~0%) and variability in CRP is high, then the power could be much greater than 90% for larger effect sizes (e.g., treatment difference = 50% and SD = 30%, effect size = 1.667).

Additional patients are being included in Phase 2 to allow for adequate estimation and improve precision of the treatment effect using the proposed Phase 3 clinical endpoint—time to improvement (2 points) in clinical status (7-point ordinal scale)—prior to making adaptations to the Phase 3 study design. The total sample size for Phase 2 is updated to be approximately 460 patients, to include patients with all baseline IL-6 levels.

As the Phase 2 data evolve, power calculations will be reassessed, and the Phase 3 endpoints may be changed and the sample size may be re-estimated. Considering the status of the current pandemic, the study will continue to enroll regardless of the number of patients until an active decision is taken to discontinue the study for efficacy, safety, or futility.

2.3. Study Plan

The Study event table is presented in Section 10.1.

3. ANALYSIS POPULATIONS

In accordance with guidance from the International Conference for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline ICH E9 Statistical Principles for Clinical Trials (ICH, 1998), the following population of analysis will be used for all statistical analyses in the Phase 2 portion of this study.

3.1. Intention-to-Treat (ITT) population

The intention-to-treat (ITT) population includes all randomized patients who received at least one dose of the study drug. Analysis of the ITT population will be done according to the initial treatment assigned to the patient (as randomized). The ITT population may be used for sensitivity analysis of the efficacy endpoints, as well as for analysis of data including (but not limited to) demographics and baseline characteristics.

3.2. Modified Intention-to-Treat (mITT) population

The modified intention-to-treat (mITT) population includes all randomized patients who received at least one dose of the study drug and have high baseline IL-6 levels (>ULN). Analysis of the mITT population will be done according to the initial treatment assigned to the patient (as randomized). The mITT population will be the primary population for analysis of primary and secondary efficacy endpoints, as well as for data including (but not limited to) demographics and baseline characteristics.

3.3. Per Protocol Set (PPS)

The per protocol population set (PPS) includes all ITT patients who did not have any relevant major protocol deviations, e.g., patients who are randomized and treated, but do not have laboratory-confirmed SARS-CoV-2 infection will be excluded from PPS. The final determination of the exclusion of patients from the PPS will be made prior to the first database lock (for Phase 3 only). Analysis of the PPS will be done according to the treatment the patient actually received (as treated). The PPS will be used for sensitivity analysis of the primary efficacy endpoint in Phase 3 portion of the study.

3.4. Safety (SAF) population

The Safety population (SAF) includes all randomized patients who received at least one dose of the study drug. Analysis of the Safety population will be done according to the treatment received (as treated). Determination of “as treated” will be based on the actual study drug received on Day 1.

3.5. Pharmacokinetics Analysis Sets

The pharmacokinetics (PK) analysis population includes all patients who received any study drug and who had at least 1 non-missing result following the first dose of study drug.

4. ANALYSIS VARIABLES

4.1. Demographics and Baseline Characteristics

Demographic and baseline characteristic variables include the following:

- Age at screening (years)
- Age group (18 to <65, 65 to <85, >=85 years)
- Sex (Male, Female)
- Race (Asian, Black/African American, Native Hawaiian/Other Pacific Islander, White and Other)
- Ethnicity (Hispanic or Latino, Not-Hispanic or Latino)
- Baseline Weight (kg)
- Baseline Height (cm)
- Baseline Body mass index (BMI) (kg/m^2) calculated from weight and height
- Severity of illness (per IRT) (severe, critical, multi-system organ dysfunction or immunocompromised [recorded on CRFs or IRT, as applicable])
- Systemic corticosteroids use (per IRT)
- SARS-CoV-2 virus result (per CRF) (Positive/Negative) (Note: SARS-CoV-2 infection is laboratory-confirmed by PCR prior to randomization)
- Cycle Threshold, if SARS-CoV-2 positive (per CRF)

Other baseline disease characteristic variables for this study population are as follows.

Pneumonia status at baseline – (selected from Pneumonia Status at Baseline CRF)

- History of or current chronic hypercapnic respiratory failure (yes/no) – (selected from NEWS2 CRF at baseline or Pneumonia status at baseline CRF, depending on original or amended versions of CRFs)
- Presence of pneumonia based on historical chest X-ray or CT scan (yes/no)
- Duration of pneumonia prior to baseline (calculated as earliest onset date for symptoms of pneumonia to first dose date)
- Presence of rales/crackles on lung auscultation (yes/no)
- Presence of documented fever in medical record (yes/no)
- Highest temperature recorded 24 hours prior to dosing
- Use of supplemental oxygen by device type (yes/no)
  - Nasal cannula
  - Face mask

  - Non-rebreather mask
  - High-flow nasal cannula
  - Other (specify)
- Use of non-invasive ventilation (yes/no)
- Use of invasive ventilation (mechanical ventilation) (yes/no)
- Use of vasopressors (yes/no)
- Use of extracorporeal life support (yes/no)
- Use of renal replacement therapy (yes/no)
- Number of patients requiring treatment in intensive care unit (ICU)

Oxygen administration and Oxygenation at baseline

- Type of oxygen delivery device
  - Nasal cannula
  - Simple face mask
  - Non-rebreather face mask,
  - High-flow nasal cannula
  - Non-invasive ventilation
  - Invasive mechanical ventilation
  - Extracorporeal life support
  - Other)
- SpO_2 % (peripheral capillary oxygen saturation) (range is 0% to 100%)
- FiO_2 (fraction of inspired oxygen) (range is 0.0 to 1.0) – See Section 4.5.3.2 for derivations
- SpO_2 / FiO_2 ratio– See Section 4.5.3.2 for derivations

Hospital or ICU stay

- Length of hospital stay including ICU prior to randomization (days)
- Admitted into ICU during hospital stay prior to randomization (yes/no)
- Length of ICU stay prior to randomization (days)

4.2. Medical History

Medical history will be coded to a Preferred Term (PT) and associated primary System Organ Class (SOC) according to Medical Dictionary for Regulatory Activities (MedDRA®) version 23.0.

4.3. Prior / Concomitant Medications or Procedures

Medications/Procedures will be recorded from the day of informed consent until the final study assessment (Day 29 or discharge or death). Medications will be coded to the ATC level 2 (therapeutic main group) and ATC level 4 (chemical/therapeutic subgroup), according to WHO Drug Dictionary (WHODD) version 202003. Patients will be counted once in all ATC categories linked to the medication.

Prior medications/procedures are: medications taken or procedures performed prior to administration of the study drug.

Concomitant medications/procedures are: medications taken or procedures performed following the first dose of study drug through the final study assessment (Day 29 or discharge or death). This includes medications taken that started before the study and are ongoing during the study.

Concomitant medications in a hospitalized population change daily and are difficult to collect in the setting of a pandemic with limited resources in some settings. It is therefore difficult to attribute to success and failure of therapy and impact on safety. Therefore, only select concomitant medications will be captured in this trial. The select list of medications include corticosteroids, remdesivir, lopinavir-ritonavir, chloroquine, hydroxychloroquine, interferon beta, and convalescent serum.

Analysis of medications data will be focused on the targeted medications (specified in the protocol) that are expected to be reviewed and recorded by sites. **Targeted medications** include but are not limited to:

- antipyretics, such as aspirin, acetaminophen, ibuprofen, and other non-steroidal anti inflammatory drugs (NSAIDs)
- warfarin
- cyclosporine A
- theophylline
- digoxin
- antiepileptics, such as carbamazepine (Carbatrol®, Tegretol®), divalproex (Depakote®), phenytoin (Dilantin®), valproic acid (Depakene®);
- antiarrhythmics, such as disopyramide (Norpace®), procainamide (Procan®, Pronestyl®), quinidine (Quinidex®, Quin Release Quin- G®)
- antivirals, such as remdesivir, lopinavir-ritonavir, ganciclovir, acyclovir, valganciclovir
- chloroquine
- hydroxychloroquine
- interferon beta
- corticosteroids

- convalescent serum
- angiotensin receptor blockers, such as Azilsartan (Edarbi), Candesartan (Atacand), Eprosartan (Teveten), Irbesartan (Avapro), Losartan (Cozaar), Olmesartan (Benicar), Telmisartan (Micardis), Valsartan (Diovan)
- angiotensin converting enzyme inhibitors: benazepril (Lotensin), captopril (Capoten), enalapril (Vasotec), fosinopril (Monopril), lisinopril (Prinivil, Zestril), moexipril (Univasc), perindopril (Aceon), quinapril (Accupril)

4.4. Rescue Medication/or Prohibited Medication During Study

There are no rescue, prohibited or permitted medications in this study, except for those in the exclusion criteria of study enrollment. Patients may continue their normal regimen of medications and procedures. All data collected on medications/procedures (pre-treatment and concomitant) will be summarized.

4.5. Efficacy Variables

4.5.1. Primary Efficacy Variable – Percent Change from Baseline in C-Reactive Protein

The primary efficacy variable is the **percent change from baseline in C-reactive protein (CRP) levels at Day 4**.

CRP levels measured by local labs at both Day 1 (baseline) and Day 4 will be used (by default) to compute the primary efficacy variable. If both local lab and central lab data are available at both time points, then central lab data will be used.

In case of missing local lab CRP data, every attempt will be made to send blood samples for testing to central labs and the following imputation rules will be applied.

1. If local lab CRP data is missing at Day 1, then central lab CRP data at both Day 1 and Day 4 will be used. If the Day 1 blood sample was not taken or not evaluable, CRP from levels from the screening period (at any time prior to Day 1) will be used.
2. If local lab data is available for Day 1 but missing for Day 4, then missing CRP levels will be imputed by local lab CRP data for Day 3 or Day 5 when available, in this order of priority.
3. If local lab CRP data at both Day 1 and post-baseline Days 3/4/5 are missing, then central lab data at both Day 1 and Day 4 will be used.
4. If local lab CRP data at Day 1 and Days 3, 4, 5 are all missing and no samples are available for testing at the central lab, then CRP data will be left missing (no imputation).

Baseline CRP levels will be defined as CRP levels measured at baseline visit (Day 1) based on blood sample taken prior to dosing.

Percent change from baseline in CRP levels at Day 4 will be calculated on both raw scale and natural log-transformed data (used to estimate treatment effect based on ANCOVA model; Section 5.7.1).

- Raw scale: $[(\text{CRP at Day 4})] - (\text{Baseline CRP})/(\text{Baseline CRP})$
- Natural-log scale: $\{[\ln(\text{CRP at Day 4}) - \ln(\text{Baseline CRP})]/\ln(\text{Baseline CRP})\}$.

4.5.2. Key Secondary Efficacy Variable – Time-to-improvement (≥2 points) in clinical status assessment using 7-point ordinal scale

The key secondary efficacy variable for Phase 2 is **time-to-improvement (≥2 points) in clinical status assessment from baseline using the 7-point ordinal scale**.

This variable will be used to evaluate the following 2 key secondary efficacy endpoints in 2 different sets of patients:

1. Time to improvement (≥2 points) in clinical status assessment from baseline on the 7-point ordinal scale in severe or critical patients with high baseline IL-6 levels (> upper limit of normal), and
2. Time to improvement (≥2 points) in clinical status assessment from baseline on the 7-point ordinal scale reporting in severe or critical patients with all IL 6 levels.

The **ordinal scale** is an assessment of the clinical status of a patient (Peterson, 2017 [4]). The 7-point ordinal scale is as follows:

1. Death;
2. Hospitalized, requiring invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO);
3. Hospitalized, requiring non-invasive ventilation or high flow oxygen devices;
4. Hospitalized, requiring supplemental oxygen;
5. Hospitalized, not requiring supplemental oxygen – requiring ongoing medical care (COVID-19 related or otherwise)
6. Hospitalized, not requiring supplemental oxygen – no longer requires ongoing medical care
7. Not hospitalized

The clinical status of a patient will be assessed in the morning to consider the worst assessments for the previous day (ie, midnight to midnight; 00:00 – 00:00 [24 hour clock]). If it is the first assessment and 24 hours of data are not available, then the clinical status will be recorded at randomization. Data for clinical status (ordinal scale) is recorded on the Clinical Status Assessment Ordinal Scale case report form (CRF).

Baseline clinical status (prior to dosing on Day 1) in this hospitalized patient population (randomized and treated) will take values 2, 3, or 4, i.e., all patients entering the study will require either supplemental oxygen or require ventilation (non-invasive, invasive or ECMO), and must be hospitalized (as specified in study Inclusion Criteria). Patients who are randomized but not treated, for example who die or are discharged prior to dosing, may have baseline value of 1 or 7, respectively. Randomized but not treated patients are excluded from the analysis.

Post-baseline clinical status can take values 1 through 7 on ordinal scale and data will be as recorded on the ordinal scale case report form (CRF). However, since the ordinal scale on a given study is recorded in the morning and it reflects the patient's clinical status during the prior 24 hour period, derivations for death and hospital discharge on that study day will be done as follows.

Handling of deaths or discharge on ordinal scale (post-baseline):

For patients who die, the clinical status ordinal scale value will be retained for the study day as recorded, but the clinical status for the next study day will be derived as "1=Death". After this study day, the observed clinical status ordinal value will be kept missing for derivation of time-to-improvement variable (described below) and this value will be carried forward until end of study. Death date may be obtained from Adverse Events or Study Completion CRFs or phone call at Day 60 (end of study).

For example, a patient's clinical status may be recorded as "2= Hospitalized, requiring invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO)" on the morning of Day 5, but the patient dies later in the evening of Day 5. In this case, clinical status (observed) = 2 for Day 5, and clinical status (derived) = 1 for Day 6 and carried forward for all subsequent study days.

Similarly, hospital discharge information may be obtained from Hospital & ICU admission & discharge CRF. Earliest available date of hospital discharge will be used and clinical status imputed to 7, for the study day after discharge date. After this study day, the observed clinical status ordinal value will be kept missing for derivation of time-to-improvement variable (described below) and this value will be carried forward until end of study. If it is learned during follow-up that some patients are subsequently readmitted for COVID-related reasons, the data will be explored to assess effects of readmission in the analyses.

Handling of other missing data on ordinal scale (post-baseline):

For patients who are alive and not discharged (i.e., still hospitalized), missing clinical status (ordinal scale) value on a given study Day X will be imputed using data on the type of oxygen delivery device used by the patient on the previous day, i.e., Day X-1. Clinical status = 2 (if using invasive mechanical ventilation or ECMO), or 3 (if using non-invasive mechanical ventilation or high-flow nasal cannula), or 4 (if using nasal cannula, simple face mask, or non-breathter face mask) or 5/6 (if not using supplemental oxygen). Clinical status of 5 or 6 will be assigned through medical review of data, or query of site.

Time-to-improvement (≥2 points) in clinical status assessment from baseline:

Patients with (post-baseline clinical status – baseline clinical status value) ≥2 will be regarded as an improvement (event). Time to improvement (event) is computed as the (first date of improvement – first dose date). Patients who do not experience improvement of 2 points or more on the ordinal scale will be censored at the last observed time point. Patients who die at any time in the study will be censored at Day 60 (end of study – phone call follow-up) in case of final database lock for Phase 2 (or censored at date of data cut-off in case of interim locks).

4.5.3. Other Secondary Efficacy Variables

The other secondary efficacy endpoints specified in the protocol are as follows. Details for deriving the variables associated with these endpoints are in Sections 4.5.3.1, 4.5.3.2, 4.5.3.3, 4.5.3.4, 4.5.3.5, 4.5.3.6, and 4.5.3.7.

1. Time to resolution of fever for at least 48 hours without antipyretics or until discharge, whichever is sooner, in patients with documented fever ≥38°C (oral), ≥38.4°C (rectal or tympanic), or ≥37.6°C (temporal or axillary) at Baseline

   (Resolution of fever is defined as postbaseline body temperature <37.2°C (oral), or <37.6°C (rectal or tympanic) or <36.8°C (temporal or axillary).
2. Time to resolution of fever (defined as above) for at least 48 hours without antipyretics or until discharge, whichever is sooner, in patients defined as above, by clinical severity.
3. Time to resolution of fever (defined as above) for at least 48 hours without antipyretics or until discharge, whichever is sooner, in patients defined as above, by baseline IL-6 levels.
4. Time to improvement in oxygenation (increase in SpO_2/FiO_2 of 50 or greater compared to the nadir SpO_2/FiO_2) for at least 48 hours or until discharge, whichever is sooner.
5. Time to improvement in oxygenation (increase in SpO_2/FiO_2 of 50 or greater compared to the nadir SpO_2/FiO_2) for at least 48 hours, or until discharge, whichever is sooner, by clinical severity
6. Time to improvement in oxygenation (increase in SpO_2/FiO_2 of 50 or greater compared to the nadir SpO_2/FiO_2) for at least 48 hours, or until discharge, whichever is sooner, by baseline IL-6 level
7. Time to resolution of fever (as defined above) and improvement in oxygenation (as defined above)
8. Mean change in the 7-point ordinal scale from baseline to Days 3, 5, 8, 11, 15, and 29 (or until discharge)
9. Percentage of patients in each clinical status category using the 7-point ordinal scale at Days 3, 5, 8, 11, 15, and 29
10. Time to discharge or to a NEWS2 of ≤2 and maintained for 24 hours, whichever occurs first
11. Change from baseline to Days 3, 5, 8, 11, 15, and 29 in NEWS2
12. Days with fever (≥38°C °C [oral], ≥38.4°C [rectal or tympanic], or ≥37.6°C [temporal or axillary])
13. Proportion of patients alive, off oxygen at Day 29
14. Days of resting respiratory rate >24 breaths/min (recorded at least once each day)
15. Days of hypoxemia (SpO_2 ≤93% on room air, or requiring supplemental oxygen, or mechanical ventilatory support)
16. Days of supplemental oxygen use

17. Time to saturation ≥94% on room air
18. Ventilator free days in the first 28 days (to Day 29)
19. Initiation of mechanical ventilation, non-invasive ventilation, or use of high flow nasal cannula (for those not requiring these interventions at baseline)
20. Admission into an ICU (among those not in an ICU at baseline)
21. Days of hospitalization among survivors
22. All-cause mortality

4.5.3.1. Fever

Body temperature data is collected on the following CRFs using various routes, i.e., oral, rectal, tympanic, temporal or axillary, or other (e.g., urinal catheter):

- Pneumonia status at baseline (highest temperature recorded within 24 hours prior to dosing)
- Body temperature (Day 1 – Day 3)
  - For baseline, pre-dose on Day 1 will be used
  - For post-baseline (after dosing on Day 1), temperature recorded in morning, noon, evening, bedtime will be used
- Body temperature (Day 4 onwards) – all post-baseline
- Body temperature – baseline or post-baseline (highest temperature during the day)
- National Early Warning Score 2 (NEWS2)
- Vital signs

Baseline Body Temperature for a patient is defined as the maximum temperature (using original route) at any time prior to dosing obtained from data across all CRFs. In order to determine the maximum, all body temperatures recorded prior to dosing with various routes (oral, rectal/tympanic, or temporal/axillary) will be first converted to oral scale:

- (Rectal or Tympanic) – 0.4 °C = Oral °C
- (Temporal or Axillary) + 0.4 °C = Oral °C

Other routes of body temperature (e.g., urinary catheter) will follow same conversion rule as rectal.

Maximum body temperature using the original route (prior to dosing) that corresponds to the maximum oral body temperature (prior to dosing) will be the Baseline Body Temperature.

Similarly, Post-baseline Body Temperature will be derived as the maximum body temperature (keeping the original route) on a given study day.

Presence of Fever (yes/no) on a given study day is defined as body temperature ≥38°C (oral), ≥38.4°C (rectal or tympanic), or ≥37.6°C (temporal or axillary).

Resolution of fever is defined as post-baseline body temperature <37.2°C (oral), or <37.6°C (rectal or tympanic) or <36.8°C (temporal or axillary), for at least 48 hours without antipyretics.

Antipyretics are defined in Appendix 10.3. Resolution of fever is defined only in patients with presence of fever at Baseline (Day 1) (fever defined as above).

Time to resolution of fever is defined as [min(earliest date of resolution of fever, hospital discharge date) – first dose date]+1, in patients who experience resolution of fever (event). Patients who do not experience resolution of fever will censored at the last observed study day. Patients who die at any time in the study will be censored at Day 60 (end of study), even if fever is resolved.

4.5.3.2. Oxygenation

All oxygenation-related data will be summarized from the Oxygen Administration and Oxygenation CRF. Supplemental oxygen may be administered to patients through an oxygen delivery device (e.g., nasal cannula, simple face mask, non-breather face mask, high-flow nasal cannula, non-invasive ventilation, invasive mechanical ventilation, extracorporeal life support, etc.).

As applicable, the following supplemental oxygen / FiO_2 data will be recorded:

- Oxygen flow rate in Liters/min (if receiving nasal cannula, simple face mask, non-rebreather mask)
- FiO_2 (if receiving high flow nasal cannula, non-invasive ventilation, invasive mechanical ventilation, or extracorporeal life support)

For devices that report both Oxygen flow rate and FiO_2, only FiO_2 data will be used. For devices reporting only oxygen flow rate, this data will be converted to FiO_2 as given in Appendix 10.3.

Resting SpO_2 (peripheral capillary oxygen saturation) will also be measured to assess arterial oxyhemoglobin saturation (ie., SpO_2 is the percentage of hemoglobin containing oxygen compared to total amount of hemoglobin in the blood). SpO_2 will be measured simultaneously with the recorded supplemental oxygen/FiO_2 data.

The variable **SpO_2** (peripheral capillary oxygen saturation) will be reported in % units (range: 0% to 100%) and **FiO_2** (fraction of inspired oxygen) will be reported in decimals (range: 0.0 to 1.0).

Baseline SpO_2 / FiO_2 ratio will be the last available ratio prior to dosing. Both SpO_2 and FiO_2 values must be nonmissing; missing baseline values will not be imputed.

Post-baseline SpO_2 / FiO_2 ratio will be the minimum of all ratios (in case of multiple values) on a given study day.

Change from baseline in SpO2 / FiO2 ratio will be calculated for each study day.

Improvement in oxygenation is defined as any post-baseline SpO_2/FiO_2 ratio $\geq$ nadir + 50, after the nadir occurs, where nadir is the lowest SpO_2/FiO_2 ratio

Time to improvement in oxygenation is defined as the time from first dose date to the first improvement in oxygenation, lasting for at least 48 hours or until discharge, whichever is sooner. Patients who do not experience improvement in oxygenation will be censored at the last observed time point. Patients who die will be censored at Day 60 (even if improvement in oxygenation is yes before they die).

4.5.3.3. Clinical status using 7-point ordinal scale

See definition and derivations in Section 4.5.2.

Number and percentage of patients in each clinical status category (ordinal scale) will be calculated for each study day, based on observed data. Handling of data for deaths, discharge and missing data is given in Section 4.5.2. Percentages will be calculated based on number of patients in a given category on that study day.

Change from baseline in clinical status (ordinal scale) per patient will be calculated for each post-baseline study day as points of change. In this study population, baseline will only take values 2, 3, 4 (on ordinal scale) and post-basline will take values 1, 2, 3, 4, 5 6, 7. Change from baseline will take values: e.g., -3, ,2, -1, 0, +1, +2, +3, +4, +5 (representing worsening, no change, or improvement). Number and percentage of patients will be calculated for each study day, for the change from baseline points, e.g., -3, -2, -1, 0, +1, +2, +3, +4 and +5.

4.5.3.4. National Early Warning Score 2 (NEWS2)

The NEWS2 scoring system is a composite score derived from 7 physiological parameters (Williams, 2019 [6]), (See Table 1):

1. Respiratory rate (per minute),
2. SpO_2 Scale 1 (%) or SpO_2 Scale 2 (%),
3. Use of air or oxygen (?),
4. Systolic blood pressure (mm Hg),
5. Pulse (per minute),
6. Consciousness, and
7. Temperature (°C).

Notes: Level of consciousness of a patient will be recorded as alert (A), confusion (C), arousable to voice (V), pain (P), or unresponsive (U). SpO_2 Scale 2 is used only in patients with history of or current Chronic hypercapnic respiratory failure (as recorded on NEWS2 CRF in the original CRFs, and on Pneumonia Status at Baseline CRF after Protocol Amendments #1 /#3). For all other patients, SpO_2 Scale 1 is used.

The NEWS2 score is assessed in patients daily in the morning from baseline visit (Day 1) until Day 29, discharge or death, whichever is sooner.

If any component (physiological parameter) is missing, then that value will be imputed using last observation carried forward.

Time to (discharge or to a NEWS2 score of ≤2) and maintained for 24 hours, whichever occurs first, will be derived. Patients who die will be censored at Day 60.

Note that the Total NEWS2 score for a patient can theoretically range from 0 to 23. For descriptive analysis, NEWS2 score will be regarded as continuous data.

Table 1: The National Early Warning Score 2 (NEWS2) Scoring System

Physiological parameter	Score 3	2	1	0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	≥25
SpO_2 Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO_2 Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51–90	91–110	111–130	≥131
Consciousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

4.5.3.5. Ventilation outcomes

Ventilator-free days based on Study Day 29

Ventilator-free days (VFD) depend on both duration of ventilation and mortality through study Day 29. The maximum number of VFD can be 28 days. (Assuming that patient is observed at the same time on each study day, there are 28 24-hour periods from Day 1 to Day 29.) VFD will be calculated and reported only in patients using invasive mechanical ventilation in Critical COVID-19 patients at baseline.

Assisted ventilation is defined as the use of invasive mechanical ventilation or extracorporeal membrane oxygenation. For those receiving assisted ventilation, the **duration of ventilation** is defined as [Last date of assisted ventilation in the hospital – *max*(first date of assisted ventilation, first dose date)], if last day is prior to Day 29. (Note: Some patients may be randomized while on assisted ventilation. Hence maximum of first date of assisted ventilation and first dose date is used.) Otherwise, **duration of ventilation** is (Date of Study Day 29 - first date of assisted ventilation). (Isolated periods of ventilation briefer than 24 hours for surgical procedures and ventilation solely for sleep disordered breathing will not count towards duration of ventilation.)

In patients who never require assisted breathing, **duration of ventilation** is zero.

Ventilator-free days are defined as 28 minus duration of ventilation, with following considerations.

- Patients discharged prior to Day 29 (but not to home) on unassisted breathing will be assumed to remain on unassisted breathing through Day 29.

- For patients who experience multiple episodes of assisted ventilation (e.g. may be on a ventilator, then come off the ventilator and be on a ventilator again) ventilator free days will be computed between each episode of ventilation and total VFD calculated.
- Patients who do not survive 29 days will be assigned zero VFD.

Number and percentage of patients with initiation of mechanical ventilation, non-invasive ventilation, or use of high flow nasal cannula (for those not requiring these interventions at baseline) will be calculated.

Ventilator-free days based on Study Days 8, 15, 22

VFD variable for each patient will also be calculated similarly (as above) based on data up to Study Day 8, 15, and 22 (replace Day 29 by respective study day). Maximum VFD will be 7 days, 14 days, and 21 days, respectively.

4.5.3.6. Other respiratory outcomes

The following respiratory outcomes will be summarized descriptively for each treatment group. Descriptive statistics will be provided for number of days related outcomes in terms of n, mean, sd, median, min and max.

- Proportion of patients who are alive and off oxygen at Day 29
  - i.e., no supplemental oxygen or mechanical ventilation used, or clinical status is 5, 6, or 7.
- Number of days of resting respiratory rate >24 breaths/min (recorded at least once each day) (obtained from NEWS2 CRF)
- Number of days of hypoxemia (SpO_2 ≤93% on room air, or requiring supplemental oxygen, or mechanical ventilatory support) (obtained from Oxygen administration and Oxygenation CRF)
- Number of days of supplemental oxygen use
- Time to saturation ≥94% on room air calculated as days until SpO_2 ≥ 94% without supplemental oxygen or ventilation

4.5.3.7. Hospitalization, Intensive Care Unit (ICU) and Mortality

Number and percentage of patients admitted into the ICU out of those not in the ICU at baseline will be summarized by treatment group. Among patients who are discharged from the hospital alive, **number of days of hospitalization** will also be summarized by treatment group using descriptive statistics. **All-cause mortality** will be summarized as proportion of patients who died. **Time-to-death** from first dose date will be analyzed using survival analysis methods. Death (mortality) data will be based on the recorded death date.

ICU-free days will be calculated using the same logic as Ventilator-free days at study days 8, 15, 22 and 29 (Section 4.5.3.5).

4.6. Safety Variables

4.6.1. Adverse Events and Serious Adverse Events

Serious adverse events and AESIs will be collected from the time of informed consent signature and then at each visit through Day 29 or discharge or death, whichever is sooner. Patients discharged prior to Day 29 will have a follow-up phone call on Day 29 to assess vital status, collect data on serious adverse event and history of hospital re-admission. At the Day 60 phone call follow-up, data on vital status of the patient (alive or dead) and history of hospital re-admission will be collected. All adverse events are to be coded to a "Preferred Term (PT)" and associated primary "System Organ Class (SOC)" according to the Medical Dictionary for Regulatory Activities (MedDRA version 23.0).

An Adverse Event is any untoward medical occurrence in a patient administered a study drug which may or may not have a causal relationship with the study drug.

A Serious Adverse Event is an adverse event (AE) that is classified as serious according to the criteria specified in the protocol.

Infusion reactions are defined as any relevant AE that occurs during the infusion or within 2 hours after the infusion is completed.

The severity of AEs will be graded using the NCI-CTCAE v5. Adverse events not listed in the NCI-CTCAE v5 will be graded according to the following scale:

Table 2: Grading System for Adverse Events Not Listed in NCI-CTCAE

Grade	Severity	Description
1	Mild	Asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
2	Moderate	Minimal, local, or noninvasive intervention indicated; limiting age-appropriate instrumental ADL*.
3	Severe	Severe or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care ADL**.
4	Life–threatening	Life-threatening consequences; urgent intervention indicated.
5	Death	Death related to AE

* Instrumental Activities of Daily Living (ADL) refer to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.

** Self-care ADL refer to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden.

Laboratory results, vital signs, or ECG (if feasible) abnormalities will be recorded as AEs if they are medically relevant: symptomatic, requiring corrective therapy, leading to treatment discontinuation and/or fulfilling a seriousness criterion.

The **secondary safety variables** are (also see Section 4.6.2):

1. Incidence of serious adverse events
2. Incidence of Grade 4 neutropenia (absolute neutrophil count (ANC) $< 500/mm^3$)
3. Incidence of severe or life-threatening bacterial, invasive fungal, or opportunistic infection
4. Incidence of severe or life-threatening bacterial, invasive fungal, or opportunistic infection in patients with Grade 4 neutropenia (ANC $< 500/mm^3$)
5. Incidence of hypersensitivity reactions
6. Incidence of infusion reactions
7. Incidence of gastrointestinal perforation
8. White cell count, hemoglobin, platelets, creatinine, total bilirubin, ALT, AST, on Days 1, 3, 5, 8, 11, 15, and 29 (if still hospitalized)

4.6.2. Adverse Events of Special Interest

Adverse events of special interest (AESI) are AEs (serious or non-serious) of scientific and medical interest specific to this drug program, for which ongoing monitoring and rapid communication by the investigator to the sponsor will be done.

In this study, the AESIs are listed below:

- Grade 4 neutropenia (absolute neutrophil count (ANC) $< 500/mm^3$) :
  - selected from Local Lab-Hematology CRF
- Grade 4 neutropenia (ANC $< 500/mm^3$) with concurrent severe or life-threatening bacterial, invasive fungal, or opportunistic infection :
  - selected from Adverse Event CRF
- Grade ≥2 infusion related reactions
  - selected from Adverse Events CRF
- Grade ≥2 hypersensitivity reactions
  - selected from Adverse Events CRF
- Increase in alanine transaminase (ALT) or aspartate aminotransferase (AST) ≥ 3xupper limit of normal (ULN) (for patients with normal baseline) or >3X ULN AND at least 2 fold increase from baseline value (for patients with abnormal baseline)
  - selected from Local Lab-Chemistry CRFs
- Invasive bacterial or fungal infections of clinical significance with confirmed diagnosis based on the investigator's assessment with appropriate diagnostic workups and consultations
  - selected from Adverse Event CRF

4.6.3. Laboratory Safety Variables

The clinical laboratory data consists of blood chemistry (including C-Reactive Protein, liver function tests, creatinine and other), hematology, urinalysis, infection testing, SARS-Cov-2 RT-PCR, serum sIL-6R, and other (as specified in the protocol).

Clinical laboratory values will be converted to standard international (SI) units and grouped by function in summary tables. Conventional unit may be provided. Functions are defined as follows:

- Liver function including ALT, AST, alkaline phosphatase, total bilirubin,
- Renal function including creatinine, uric acid,
- Electrolytes including sodium, potassium,
- C-Reactive Protein (CRP),
- Creatine Phosphokinase (CPK)
- Metabolic parameters including total proteins, albumin,
- White blood cells (WBCs) including WBCs count and differential count (neutrophils, lymphocytes, eosinophils, basophils, monocytes),
- Red blood cells (RBCs) and platelets including red blood cells count, hemoglobin, hematocrit and platelets count,
- Other

4.6.4. Vital Signs

Vital signs, including blood pressure, pulse, and respiration, are recorded at multiple time points according to Schedule of Time and Events table (See Section 10.1). Temperature is also recorded at multiple time points (See Section 4.5.3.1).

4.6.5. Physical Examination Variables

A targeted physical examination including lung auscultation will be performed at time point according to Schedule of Time and Events table (See Section 10.1). Care will betaken by the investigator to examine and assess any abnormalities that may be present, as indicated by the patient's medical history.

During screening period, if any existing clinically significant abnormalities are present, these will be recorded in the Medical History CRF. Post-screening period, if any new clinically significant abnormalities are present (per investigator discretion), the relevant event will be recorded in the Adverse Event CRF, if applicable

4.7. Pharmacokinetic Variables

The PK variable is the concentration of sarilumab and sIL-6R in serum at each time point specified in the Schedule of Time and Events table (See Section 10.1) .

4.8. Pharmacodynamic and Other Biomarker Variables

Exploratory endpoint variables include measurement of SARS-CoV-2 in OP or NP swabs over time using RT-PCR. Qualitative (positive or negative) or relative quantitation of viral copies may be evaluated. Pharmacodynamic variables may include the time to reach a negative OP or NP RT PCR result.

Additional biomarker testing may include, but not be limited to, evaluation of inflammatory cytokines in serum, and ANC.

Pharmacodynamic variables may include the time to nadir (or peak), descriptive statistics of absolute value, absolute change from baseline, percent change from baseline by nominal time (visit), and area under the curve (AUC) of mean and median change from baseline for IL-6 and ANC.

These results may be reported outside the clinical study report (CSR).

5. STATISTICAL METHODS

For continuous variables, descriptive statistics will include the following: the number of patients reflected in the calculation (n), mean, median, standard deviation (sd), minimum, and maximum.

For categorical or ordinal data, frequencies and percentages will be displayed for each category.

5.1. Demographics and Baseline Characteristics

Demographic and baseline characteristics variables given in Section 4.1 will be summarized descriptively by treatment group, and all groups combined. These will be analyzed for both mITT and ITT populations.

5.2. Medical History

Medical history will be summarized by SOC and PT and by treatment group and all groups combined in the ITT population.

5.3. Prior / Concomitant Medications or Procedures

Prior or concomitant medications/procedures will be summarized by treatment groups. Focus of the results will be on the list of targeted medications (Section 4.3) in the ITT population.

5.4. Rescue/ Prohibited Medications if applicable

Not applicable. See Section 4.3.

5.5. Subject Disposition

The following summaries will be provided for both mITT and ITT populations.

- The total number of screened patients: met the inclusion criteria regarding the target indication and signed the ICF
- The total number of randomized patients: received a randomization number
- The total number of patients who discontinued the study, and the reasons for discontinuation
- The total number of patients who discontinued from study treatment, and the reasons for discontinuation.

  (Note: Until Protocol Amendment #3 was effective, patients were given only a single dose. After Protocol Amendment #4, repeat dosing is permitted after 24 hours of dosing and weekly under protocol-defined criteria. Discontinuation from study treatment refers to either study treatment interruption (if single dose) or study drug discontinuation (if multiple doses).)
- A listing of patients treated but not randomized, patients randomized but not treated, and patients randomized but not treated as randomized
- A listing of patients prematurely discontinued from treatment, along with reasons for discontinuation

(Note: Applicable only on data after Protocol Amendment #4 is in effect, as described above.)

- The ITT population and mITT population (defined in Sections 3.1 and 3.2)
- The Safety population (defined in Section 3.4)

5.6. Extent of Study Treatment Exposure and Compliance

Until Protocol Amendment #3 was effective, treated patients were given a single dose of study drug (sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo IV). After Protocol Amendment #4, repeat dosing of study drug is permitted after 24 hours if no clinical response is observed and repeat weekly dosing is permitted for patients requiring supplemental oxygen (under protocol-defined criteria).

5.6.1. Measurement of Compliance

Treatment compliance in a given patient is defined as the number of fully completed infusions of study drug divided by number of doses administered (applicable, both, to patients receiving only single dose or multiple doses since Protocol Amendment #4). Treatment compliance will be summarized by treatment group using descriptive statistics based on the Safety population.

5.6.2. Exposure to Investigational Product

Exposure to study drug will be examined for each patient as recorded on the Study Drug Administration-IV CRF. The following variables will be analyzed by treatment group:

- Duration of intravenous infusion
- Location of drug administration
- Total volume of drug administered (units: mL)
- Number of patients with total planned dose administered (yes/no)
  - If no, reason for not administration of total planned dose (equipment failure, adverse event, other)
- Number of patients with infusion interruptions
- Number of patients receiving 1, 2, 3, etc. doses
- Number of patients receving drug by time, e.g., Day 1, Day 2, and weekly thereafter (Days 8, 15, 22)

5.7. Analyses of Efficacy Variables

For the Phase 2 portion of the study, the hypothesis to be tested is the superiority of sarilumab 400 mg IV versus placebo with respect to the primary endpoint of percent change from baseline in CRP levels at Day 4 (Section 4.5.1).

The statistical hypotheses (null H_0 vs. alternative H_1) for the primary endpoint are:

$$H_0: \mu_1 = \mu_2 \quad \text{versus} \quad H_1: \mu_1 \neq \mu_2$$

where μ_1 and μ_2 are mean percent changes from baseline in CRP-levels at Day 4 (based on natural logarithm scale) in the two treatment groups.

The statistical hypotheses (null H_0 vs. alternative H_1) for the Phase 2 key secondary efficacy endpoint of time-to-improvement (≥ 2 points) in clinical status from baseline using the 7-point ordinal scale (Section 4.5.2), and other secondary endpoints of time-to-resolution of fever and time-to-improvement in oxygenation (Section 4.5.2) are stated below.

$$H_0: S_1(t) = S_2(t) \text{ for all time } t \quad \text{versus} \quad H_1: S_1(t) \neq S_2(t) \text{ for some time } t$$

where $S_1(t)$ and $S_2(t)$ are the survival probability functions of the above endpoints.

5.7.1. Analysis of Primary Efficacy Variable

The primary efficacy analysis (for Phase 2) will be a pairwise comparison between sarilumab 400 mg IV and placebo with respect to the primary endpoint of *percent change from baseline in CRP levels (natural logarithm scale) at Day 4* (Section 4.5.1) in COVID-19 patients with high baseline IL-6 levels in all disease severity strata (mITT; all disease severity strata).

CRP levels measured by local labs at both Day 1 (baseline) and Day 4 will be used (by default) to compute the primary efficacy variable. Handling of missing CRP data is described in Section 4.5.1.

Hypothesis test of superiority of sarilumab versus placebo will be based on an ANCOVA model with change from baseline in CRP levels at Day 4 (CRP levels on log-scale) as dependent variable; baseline CRP (log-scale) as covariate and treatment group, severity of illness and systemic corticosteroid use as fixed effects.

ANCOVA model is stated as below:

$$[\ln(\text{CRP at Day 4}) - \ln(\text{Baseline CRP})] = \beta * \ln(\text{Baseline CRP}) + \alpha_i + \beta_j + \delta_k$$

where $[\ln(\text{CRP at Day 4}) - \ln(\text{Baseline CRP})]$ is the dependent variable, $\ln(\text{Baseline CRP})$ is the covariate with β as the regression coefficient, and $\alpha_i, \beta_j, \delta_k$ represent the fixed effects for treatment group (i=1,2), disease severity strata (j=1,2,3), and systemic corticosteroid use (k=1,2), respectively. If the sample size in the cells for systemic corticosteroids use (strata variable) are approximately 10 or below, then this variable will be dropped from the model.

Adjusted treatment mean estimates with standard errors, the adjusted estimate of treatment mean difference with standard error, and a 95 % confidence interval for the treatment mean difference will be estimated from the model.

Treatment effect will be estimated using the adjusted treatment means (Least Squares means) from the ANCOVA model as follows:

- LS-means estimate for each treatment group= $[\ln(\text{CRP at Day 4})\widehat{\ } - \ln(\text{Baseline CRP})]$, i.e., $\ln(\text{CRP at Day}\widehat{\ 4/}\text{Baseline CRP})$
- Percent change from baseline in CRP (on raw scale) will be estimated by back transformation of the log-scale, i.e., anti-log of LS-means estimate minus 1 $\exp[\ln(\text{CRP at Day}\widehat{\ 4}/\text{Baseline CRP})] - 1$
- Treatment difference based on the LS-means difference (as described above) and corresponding 95% confidence intervals will be reported

P-values for the treatment effect from this model will be compared to the 0.05 (two-sided) level of significance.

To determine the success of the Phase 2 portion of the trial, first the data in patients with high baseline IL-6 will be analyzed (mITT; all disease severity strata). If the treatment difference is statistically significant, similar analysis will be done between sarilumab 400 mg IV and placebo on the data for the full Phase 2 ITT population, ie., without regard to baseline IL-6 levels (ITT; all disease severity strata).

Sensitivity analyses:

Missing data derivation rules described in Section 4.5.1 will be followed. Although efforts will be made to obtain data from sites (hospitals) for baseline CRP and at Day 4 during routine local lab testing, or through central labs using collected blood samples, certain CRP data may still be missing in case of some patients given the serious nature of the COVID-19 pandemic and the intense workload on the hospital health care system.

Sensitivity analysis #1

Only patients with baseline CRP data will be used in the sensitivity analysis (and those with missing baseline will be excluded). If there are any post-baseline missing data (i.e., any post-baseline through Day 5 are missing), they will be imputed assuming data are missing at random as follows.

- Missing post-baseline CRP data will be imputed using a regression method with adjustment for disease severity strata, treatment group, and baseline CRP.
- Multiple imputation method will be used to create 50 complete datasets under monotone missing data pattern assumption using PROC MI step in SAS (with seed number 6882040).
- Each imputed dataset will be analyzed using ANCOVA model as described above.
- Results from the 50 analyses will then be combined using PROC MIANALYZE step in SAS to provide the statistical inferences (treatment effect estimates, confidence intervals and p-values)

Sensitivity analysis #2

Sensitivity analysis will also be conducted similar to #1, using normalized CRP data, i.e., value/ULN to account for multiple local labs.

Sensitivity analysis #3

Descriptive analysis will be presented for Percent change from baseline for CRP based on both raw-scale and log-scale based on all available data.

Supportive analyses:

Comparison between sarilumab 200 mg IV and placebo will also be performed descriptively.

If the sarilumab 400 mg treatment arm is dropped for safety reasons, then sarilumab 200 mg will be formally compared for efficacy against placebo. Due to the pandemic nature of the COVID-19 disease with public health significance, it also possible that both sarilumab dose arms may be found to be efficacious and no multiplicity adjustment is planned in that case.

Pooled analysis (combining sarilumab 200 mg IV and 400 mg IV treatment groups) versus placebo will be also conducted using the same statistical methods as primary efficacy analysis.

Subgroup analyses:

Descriptive analyses will be performed on the primary endpoint to summarize the treatment effects by the following variables:

- Severity of illness strata at enrollment (severe, critical, multi-system organ dysfunction, immunocompromised)
- Systemic corticosteroids use (yes/no), provided cell size is larger than approximately 10
- Median duration of pneumonia at baseline
- Baseline IL-6 category (eg., ≥ or < median, tertiles, etc.)

5.7.2. Analysis of Secondary Efficacy Variables

The secondary efficacy analysis (for Phase 2) will be a pairwise comparison between sarilumab 400 mg IV and placebo for the key secondary endpoint of t*ime-to-improvement (≥2 points) in clinical status assessment from baseline using the 7-point ordinal scale* (Section 4.5.2) in

1. COVID-19 patients with high baseline IL-6 within the severe and critical strata (mITT; severe and critical strata), and
2. COVID-19 patients with all baseline IL-6 levels within the severe and critical strata (ITT; severe and critical strata).

For descriptive purpose, hypothesis test of superiority of sarilumab versus placebo will be done using the stratified log-rank test with severity of illness and systemic corticosteroid use as stratification factors. (Systemic corticosteroid use strata variable may be dropped from the model if the cell size is approximately 10 or lower, as stated in Section 5.7.1).

Estimation of treatment effect will be provided as difference in median times-to-improvement (2 points) in clinical status (placebo – sarilumab) using Kaplan-Meier survival method as well as hazard ratio using Cox proportional hazards model along with two-sided confidence intervals.

In case of non-proportional hazards, stratified weighted log-rank test may be conducted using the Fleming-Harrington family of parameters for early, middle or late differences.

P-values and 95% confidence levels will be reported, for descriptive purpose on the Phase 2 data.

Supportive analysis:

The clinical status assessment is reported using the 7-point ordinal scale (k=1,…7; 1=Death, 7=Not hospitalized; see Section 4.5.2).

This ordinal data will also be analyzed using the proportional odds model stated as below:

$$\text{logit}(\theta_{hik}) = \alpha_k + \boldsymbol{x'_{hi}\beta},$$

for stratum level h ; treatment group $i = 1,2$; and ordinal scale category $k = 2, \cdots, 7$.

Here:

The dependent variable, $\text{logit}(\theta_{hik}) = \log[\theta_{hik}/(1-\theta_{hik})]$, is the log odds of at least as favorable outcome(s) (as k or greater) over less favorable outcome(s) than k,

θ_{hik} is the cumulative probability of at least as favorable outcome as k,

$(1-\theta_{hik})$ is the cumulative probability of outcomes less favorable than k

α_k is the intercept for k^{th} level in ordinal scale,

$\boldsymbol{x}_{hi}$ is the regression variable for disease severity strata and treatment group, and

$\boldsymbol{\beta}$ is the regression coefficient which is assumed to be constant for all k levels in ordinal scale due to the proportional odds assumption.

Analysis time for clinical status based on ordinal data will include all data available up to the data lock point. Analysis population will be mITT in patients in the severe and critical strata, as well as ITT in the severe and critical strata.

Treatment effect will be estimated in terms of odds ratio (OR) for sarilumab 400 mg IV over placebo based on this model, along with 95% confidence interval for the OR. P-values for this OR will be reported for descriptive purpose. Similar analysis will be conducted for sarilumab 200 mg IV versus placebo.

Subgroup analyses:

Descriptive analyses will also be performed on the key secondary endpoint to summarize the treatment effects within the severity of illness subgroups (severe, critical, multi-system organ dysfunction, immunocompromised) by the following variables:

- Demographics (e.g., Age, gender, race)
- Comorbidities (e.g., hypertension, diabetes, obesity)
- Median duration of pneumonia at baseline
- Baseline IL-6 category (eg., ≥ or < median, tertiles, etc.)

5.7.3. Analysis of Other Secondary Efficacy

Other secondary efficacy variables listed in Section 4.5.2 will be analyzed as follows.

1. Differences in time-to-event endpoints by treatment (eg., all-cause mortality, time to resolution of fever, improvement in oxygentation, and other time-to-event endpoints) will be summarized with Kaplan-Meier estimates and 95% confidence intervals.
2. Changes from baseline in ordinal scale at all study days will be summarized by number and percentage of patients over time (eg., patients who have a 1-, 2-, 3-, 4- or 5-point improvement or no change (0) or 1-, 2-, 3-, 4-point worsening; see Section 4.5.3.3).
3. Total NEWS2 score and changes from baseline NEWS2 score at all study days will be summarized with descriptive statistics as continuous variables
4. Duration of event (eg., duration of mechanical ventilation, ventilator free days, ICU-free days, days of supplemental oxygen use etc.) will be summarized with descriptive statistics (n, mean, sd, median, quartiles, min, max) across all disease severity strata as well as by severity.
5. Incidence data (eg., patients alive and off oxygen use) will be summarized as proportions with 95% confidence intervals.
6. Continuous endpoints and their changes from baseline will be summarized through descriptive statistics including mean, standard deviation, median and quartiles.
7. Categorical data (eg., 28-day mortality or ordinal scale by day) will be summarized according to proportions.

Win ratio method (Pocock et al., 2012 [5]) on a hierarchical composite of the selected efficacy endpoints (e.g., 28-day mortality, ventilator-free days, and improvement (≥ 2 points) in clinical status based on ordinal scale) as an exploratory analysis may also be conducted.

5.7.4. Adjustment for Multiple Comparison

For the Phase 2 portion of the study, hypothesis test will be conducted for the Phase 2 primary endpoint of percent change from baseline in CRP levels at Day 4 in patients across all disease severity strata in mITT population (i.e., high baseline IL-6). If there is substantial overlap between mITT, e.g., ≥ 95%, then ITT will be the primary population. Otherwise, ITT population will be tested sequentially.

The key secondary endpoint of time-to-improvement (≥2 points) in clinical status assessment will only be descriptively analyzed. P-values will be reported for descriptive purpose.

Overall Type 1 error will be controlled at 0.05 level (2-sided) in Phase 2.

5.8. Analysis of Safety Data

The analysis of safety data will be performed on the SAF, as defined in Section 3.4.

The safety analysis will be based on the reported SAEs and AESIs and other safety information (clinical laboratory evaluations and vital signs).

Thresholds for Potential Clinically Significant Values (PCSV) in laboratory variables, vital signs and ECG are defined in Section 10.2.

The summary of safety results will be presented for each treatment group.

5.8.1. Adverse Events

The verbatim text, the PT, and the primary SOC will be listed in subject listings. Summaries that include frequencies and proportions of patients reporting AEs will include the PTs and the SOCs.

Period of observation: The observation period will be the on-treatment period defined as the day from first dose of study drug (Day 1) to study Day 29. (Patients who are discharged prior to Day 29 will receive a follow-up phone call to collect data on SAEs (if any), survival and history of hospital re-admission.)

Treatment-emergent AEs (TEAEs) are defined as those AEs that are not present at baseline or represent the exacerbation of a pre-existing condition during the on-treatment period. In this study, only SAEs and AESIs are collected. As such TEAEs will reflect this data.

For details on handling missing data and partial dates, see Section 6.

Summaries of AE incidence in each treatment group will include:

- Overview of TEAEs, summarizing number of events, summarizing number and percentage of patients within the specified category
  - Total number of TEAEs, SAEs, AESIs, serious AESIs
  - Patients with any TEAEs, any SAEs, any AESIs, serious AESIs
  - Patients with any TEAEs leading to study drug interruption or study discontinuation, leading to death
- TEAEs by system organ class (SOC) and preferred terms (PT)
  - All TEAEs
  - TEAEs by relationship to treatment (related, not related)
  - TEAEs by CTC grade (according to the grading scale outlined in Section 4.6.1), presented by SOC and PT
  - TEAEs leading to study drug interruption or study discontinuation
  - TEAEs leading to death
- AESIs by PT
- SAEs by SOC and PT

- Deaths

Counts will be provided according to treatment group for each PT within each SOC. Percentages will be calculated using the number of patients from the safety population in each treatment group.

Primary SOCs will be sorted according to decreasing order of frequency. Within each primary SOC, PTs will be sorted by decreasing frequency of investigational product.

A second type of table with counts of each PT in decreasing order of frequency will also be provided.

5.8.2. Analysis of Adverse Events of Special Interest

Summaries of AESIs (given in Section 4.6.2) will include the following and presented by PT in each treatment group:

1. Incidence of Grade 4 neutropenia (ANC < 500/mm^3)
2. Incidence of severe or life-threatening bacterial, invasive fungal, or opportunistic infection (presented in listings and summarized through patient narratives)
3. Incidence of Grade 4 neutropenia (ANC < 500/mm^3) with concurrent severe or life-threatening bacterial, invasive fungal, or opportunistic infection
4. Incidence of Grade ≥2 hypersensitivity reactions, Grade ≥ 2 infusion reactions, gastrointestinal perforation
5. White cell count, hemoglobin, platelets, creatinine, total bilirubin, ALT, AST, on all study days (available data)
6. Increase in alanine transaminase (ALT) or aspartate aminotransferase (AST) ≥ 3xupper limit of normal (ULN) (for patients with normal baseline) or >3X ULN AND at least 2 fold increase from baseline value (for patients with abnormal baseline)
7. Incidence of any invasive bacterial or fungal infections

5.8.3. Clinical Laboratory Measurements

A treatment-emergent Potential Clinically Significant abnormal value (PCSV) is a laboratory value that was normal at Screening and Baseline but abnormal after treatment with study drug, or a laboratory value that was abnormal at Baseline and exacerbates after treatment with study drug. "Exacerbations" will be identified by the Medical Monitor using clinical judgment. See Appendix 10.2 for the criteria of PCSV values. Treatment-Emergent Potentially clinically significant values (PCSVs) will be summarized by treatment group.

Baseline clinical laboratory analytes and change from Baseline in clinical laboratory analytes to each scheduled assessment time will be summarized with descriptive statistics. Summary statistics will include the number of patients, mean, median, standard deviation, quartiles, minimum, and maximum.

Listings will be provided with flags indicating out of laboratory range values.

5.8.4. Analysis of Vital Signs

Vital signs (including temperature, blood pressure, pulse, and respiration) will be summarized by Baseline and change from Baseline to each scheduled assessment time with descriptive statistics. The graphs of mean (or median) value of some vital sign parameter vs. visit will also be plotted.

5.8.5. Physical Exams

As physical examination is limited in this study, only the targeted examination of lung auscultation will be provided in patient listing.

5.9. Analysis of Pharmacokinetics, Pharmacodynamics and Biomarker Data

5.9.1. Analysis of Drug Concentration Data

The concentrations of sarilumab and sIL-6R over time and selected PK parameters, as appropriate, will be summarized using descriptive statistics.

No formal statistical hypothesis testing will be performed.

5.9.2. Analysis of Pharmacodynamic and Exploratory Biomarker Data

The concentrations of exploratory PD/Biomarkers over time will be summarized using descriptive statistics.

6. DATA CONVENTIONS

The following analysis conventions will be used in the statistical analysis.

6.1. Definition of Baseline for Efficacy/Safety Variables

Definitions of baseline for efficacy variables are defined in Section 4.5 .

For safety variables, baseline will be the latest available valid measurement taken prior to the administration of study drug.

6.2. Data Handling Convention for Efficacy Variables

6.3. Data Handling Convention for Missing Data

Rules for handling missing data for primary and secondary efficacy variables are described in Sections 4.5.1, 4.5.2, 4.5.3, 5.7.1, 5.7.2 and 5.7.3.

For other continuous variables not mentioned in above sections, missing clinical efficacy data will be imputed using last observation carried forward (LOCF) as these are hospitalized patients and missing data is assumed to be missing at random.

For categorical variables, patients with missing data will be included in calculations of percentages. Number of patients with missing data will be presented.

Handling of Medications with missing/partial dates

To determine whether a medication is prior or concomitant medication, the missing medication start date is estimated as early as possible up to first dose date, and the missing medication end date is estimated as late as possible up to Day 29. If the medication start date is missing, the onset day will not be imputed in medication listings.

Handling of Adverse events Severity and Relatedness

If the intensity of a SAE and AESI is missing, it will be classified as “severe” in the frequency tables by intensity of SAE and AESIs. If the assessment of relationship of the investigational product is missing, it will be classified as related to the investigational product.

Date of injections

Date of injection is the non-missing administration date filled in the Study Drug Administration-IV CRF. If the first dose of study drug administration date is missing (even after site is queried), then the dosing date is will be imputed with the randomization date. If any subsequent study drug administration date is missing, the date of dispensation of study drug from IRT will be used.

6.4. Visit Windows

Following windows will be used for summarizing laboratory parameters. However, for the primary efficacy analysis of CRP, no windows will be applied (See Section 5.7.1).

Table 3: Time Window for Summary of Laboratory Parameters

Visit label	Target Day	Hematology, Chemistry
Baseline	1	≤1
Day 4	4	2-5
Day 7	7	6-10
Day 15	15	11-17
Day 21	21	18-24
Day 29	29	≥25

6.5. Unscheduled Assessments

The determination of baselines and values at the end of treatment for both efficacy and safety variables will be based on scheduled available assessments and unscheduled available assessments.

Extra assessments (laboratory data or vital signs associated with non-protocol clinical visits or obtained in the course of investigating or managing adverse events) will be included in listings, but not summaries except for the endpoint determination. If more than one laboratory value is available for a given visit, the first observation will be used in summaries and all observations will be presented in listings.

6.6. Pooling of Centers for Statistical Analyses

Additional analyses on primary and key secondary efficacy endpoints may be performed by sites and pooling sites enrolling less than 5 patients.

7. INTERIM ANALYSIS

To maximize safety, approximately the first 12 patients who are randomized and receive study drug will be monitored for safety (assuming a total of ~10 patients are dosed with sarilumab 200 mg IV or 400 mg IV). Safety data on these 12 patients will be reviewed by an IDMC after the last of these patients dosed reaches Day 7. Data on the first ~12, ~25, and ~50 patients (regardless of severity of illness) who reach at least Day 7 will be reviewed by the independent data monitoring committee (IDMC) and Regeneron senior physicians to determine if a treatment arm should be discontinued for safety reasons. Selected treatment groups in the Phase 2 portion may be continued forward in the remainder of the study. In addition to these specific assessment times, the IDMC will actively monitor interim data to make recommendations about early study closure or changes to study arms throughout the course of the study.

A first-step analysis of the data will be conducted for the Phase 2 portion of the study once ~460 patients are randomized and treated for 14 days, to allow for study design adaptation of Phase 3.

Decisions on sample size re-estimation, if needed, for Phase 3 and choice of primary endpoint for Phase 3 will be based on the unblined data from Phase 2.

Dropping of a treatment arm in Phase 3 may not be feasible or needed for this single-dose study due to the rapid enrollment given the dire COVID-19 pandemic situation. However, this adaptation decision may be based on the unblinded Phase 2 data or based on recommendations from the Independent Data Monitoring Committee (IDMC).

8. SOFTWARE

All analyses will be done using SAS Version 9.4.

9. REFERENCES

1. ICH. (1996, July 30). ICH Harmonized tripartite guideline: Structure and content of clinical study reports (E3). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.
2. ICH. (1997, July 17). ICH Harmonized tripartite guideline: General considerations for clinical trials (E8). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.
3. ICH. (1998, February 5). ICH Harmonized tripartite guideline: Statistical principles for clinical trials (E9). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.
4. Peterson RL, Vock DM, Powers JH, Emery S, Cruz EF, Hunsberger S, et al. Analysis of an ordinal endpoint for use in evaluating treatments for severe influenza requiring hospitalization. Clin Trials 2017; 14(3):264-76.
5. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. European Heart Journal 2012; 33: 176-182.
6. Williams B. The National Early Warning Score 2 (NEWS2) in patients with hypercapnic respiratory failure. Clin Med (Lond) 2019; 19(1):94-5.
7. Xu X, Han M, Li T, Sun W, Wang D, Fu B, Zhou Y et al. "Effective treatment of severe COVID-19 patients with Tocilizumab." ChinaXiv preprint (2020). https://www.ser.es/wp-content/uploads/2020/03/TCZ-and-COVID-19.pdf

10. APPENDIX

10.1. Schedule of Time and Events

Study Procedure	**Screening Visit[1]**	**Baseline Visit[1]**	**Daily Follow-up Until Hospital Discharge**					**EOS[2]**
Day	**-1 or 1**	**1**	**2**	**4**	**7**	**8 to 23**	**29 or discharge[3,4]**	**60**
Window			±6 hours				±7 days	±7 days
Screening/Baseline:								
Inclusion/Exclusion	X							
Informed Consent	X							
[redacted]	X							
Demographics and Medical History[1]	X							
Randomization		X						
Treatment:								
Study Drug Administration		X	X[5]			X[6]		
			If no clinical response			Repeat weekly for patients requiring any supplemental O_2		
Assessments:								
Oxygen administration (FiO_2) and Oxygenation (SpO_2)[7]	X	X	2 times a day					
Clinical Status Assessment (7-point ordinal scale)[8]		X	Daily until discharge					
NEWS2 Score								
Air or oxygen		X	Daily in the morning until discharge					
Respiratory rate		X	Daily in the morning until discharge					
BP		X	Daily in the morning until discharge					
Pulse		X	Daily in the morning until discharge					
Consciousness		X	Daily in the morning until discharge					
Body temperature before antipyretics or 4 hours after antipyretics[9]		X	Daily in the morning until discharge					

Study Procedure	**Screening Visit[1]**	**Baseline Visit[1]**	**Daily Follow-up Until Hospital Discharge**					**EOS[2]**
Day	**-1 or 1**	**1**	**2**	**4**	**7**	**8 to 23**	**29 or discharge[3,4]**	**60**
Window			±6 hours				±7 days	±7 days
Imaging, microbiology results, and arterial blood gas results (as available)[10]			If available in the medical record					
Limited physical examination (lung auscultation only)	X							
Electrocardiogram (ECG), if feasible[11]	X							
Record Targeted Medications[12]			Daily until discharge					
Adverse Events[13]	X		X					
Pregnancy Test (WOCBP)[14]	X							
Follow-up Phone Call							X	X
Laboratory Testing:								
C-Reactive Protein (**mandatory**)	X	X		X	X			
			Add CRP to all routine labs and obtain at the same time Record all results available in medical chart					
Hematology[15]	X		Record all results available in medical chart[1] Must be collected within 48 hours prior to redosing					
Blood chemistry (including LFTs and creatinine)[16]	X		Record all results available in medical chart[1] Must be collected within 48 hours prior to redosing					
Blood cultures for bacteria and fungi (**mandatory**)[17]					X	X (**Day 15**)		
PK/Biomarkers/Research (defer to footnotes for sampling requirements):								
	-1 or 1	**1**	**2**	**4**	**7**	**8 to 23**	**29 or discharge[3,4]**	**60**
Serum for PK/Sarilumab Concentration[18]		X[1]	X[5,18]	X	X	X	X[16]	
Serum sIL-6R plus research[18]		X[1]		X	X	X	X	
Serum cytokines including IL-6 and biomarker testing[1,18]		X[1]		X	X	X	X	
Blood for PCR SARS-CoV-2[19]	X	X		X			X	
Oropharyngeal or nasopharyngeal swab for SARS-COV-2 detection and sequencing[19,20]	X	X		X			X	

Study Procedure	Screening Visit[1]	Baseline Visit[1]	Daily Follow-up Until Hospital Discharge					EOS[2]
Day	-1 or 1	1	2	4	7	8 to 23	29 or discharge[3,4]	60
Window			±6 hours				±7 days	±7 days
Blood for research plasma[19]		X					X	
[redacted]								
[redacted]		X						

1. Screening and baseline may occur on the same day. Assessments that are noted for both visits should only be assessed once. Medical history should include collecting onset of pneumonia symptoms. Body temperature, SpO_2, and FiO_2 must be collected at randomization. All samples should be collected before study drug administration at baseline visit except post-infusion PK and sIL-6R samples.
2. Patients will have an end of study (EOS) assessment to collect data on survival and history of hospital re-admission. This assessment may be performed by phone.
3. Patients discharged prior to Day 29 will have a follow-up phone call on Day 29 to collect data on survival and history of hospital re-admission.
4. Patients discharged prior to Day 29 should have a sample collected for on or before day of discharge. If day of discharge is not Day 29 and coincides with another visit, the Day 29 assessments should be performed.
5. Patients can be re-dosed at 24 hours (±6 hours) if the patient:
    a. Remains febrile OR
    b. Fails to improve gas exchange (eg, as measured by ventilator settings or O_2 requirements) OR
    c. Is hemodynamically unstable OR
    d. Exhibits other objective evidence of clinical worsening (eg, mental status change, etc.)

    After informed consent is obtained, patients enrolled prior to Amendment 4 are permitted to receive a second dose prior to commencing weekly dosing. The second dose should not be administered within 18 hours of the first dose.
6. Study drug should not be administered if the ANC is less than 500/mm^3 or ALT is >5x ULN within 48 hours of redosing. Weekly dosing begins as early as Day 8 then weekly thereafter. If a patient received a 2nd dose other than Day 2, adjust subsequent weekly

doses accordingly. Study drug should be held if there is a high degree of suspicion of active bacterial or fungal infection. A maximum of 6 doses can be administered.

7. Oxygen administration and oxygenation: refer to Section 9.2.2.3 of the protocol for details. SpO_2 must be measured after 5 minutes of rest (sitting or supine) and must be measured simultaneously with oxygen administration and ventilation data. Record oxygen flow rate (L/min) for patients receiving nasal cannula, simple face mask, or non-rebreather mask. Record FiO2 for patients receiving high flow nasal cannula, non-invasive ventilation, mechanical ventilation, or extracorporeal membrane oxygenation.
8. Clinical Status Assessment using the 7-point ordinal scale: refer to Section 9.2.2.4 of the protocol for details. The Ordinal Scale should be assessed in the morning to consider the worst assessments for the previous day (ie, midnight to midnight; 00:00 – 00:00 [24-hour clock]). If it is the first assessment and 24 hours of data are not available, report status at randomization.
9. Temperature may be measured using the following methods: oral, rectal, tympanic, or temporal according to local hospital protocols and according to the manufacturer's instructions for use of the device. Body temperature should be measured using the same method each time. Temperature should be measured predose after at least 5 minutes of rest (supine or sitting).
10. If available in the medical record, chest CT images will be collected as part of a separate effort related to this study for predictive exploratory analysis and may be provided in a separate study report.
11. ECG only if feasible. Historical ECG from current hospital admission is acceptable.
12. Targeted medications: refer to Section 9.2.3.4 of the protocol for details.
13. Adverse events: Only SAEs and AESIs will be recorded in eCRF.
14. Pregnancy testing to be performed in women of childbearing potential (WOCBP) only. Serum or urine pregnancy test are both acceptable.
15. Hematology: refer to Section 9.2.3.5 of the protocol for details. CBC is required prior to randomization (standard of care labs may be used). After Day 1, CBC will not be performed as a study procedure. When CBC is performed as part of the patient's clinical care, the results will be entered in eCRF.
16. Blood Chemistry: refer to Section 9.2.3.5 of the protocol for details. LFTs and creatinine are required prior to randomization (standard of care labs may be used). After Day 1, LFTs and creatinine will not be performed as a study procedure. When chemistries are performed as part of the patient's clinical care, the results will be entered in eCRF.
17. If patient is discharged before Day 15, obtain blood culture for bacteria and fungi prior to discharge.
18. Sample collection for serum:

- Samples collected on dosing days are mandatory:
  - One predose (as close to initiation of treatment as reasonable) and
  - One within 60 minutes after the end-of-infusion (EOI). The EOI sample or flush should be collected from the arm, contralateral to that used for IV infusion, if possible. If not medically feasible, the sample can be drawn from the same arm
- Day 4 samples are mandatory (if PPE and appropriate lab facilities are available).
- The Day 1 predose sample and Day 29 or Early Termination PK sample may be used for ADA analysis.

Day 1/baseline samples are mandatory.

19. Swab and tests will be for exploratory analysis only not for inclusion or diagnosis

[redacted]

10.2. Criteria for Potentially Clinically Significant Values (PCSV)

Parameter	PCSV	Comments
Clinical Chemistry		
ALT*	>3 and ≤ 5 ULN and baseline ≤ 3 ULN* >5 and ≤ 10 ULN and baseline ≤ 5 ULN >10 and ≤ 20 ULN and baseline ≤ 10 ULN >20 ULN and baseline ≤ 20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Each category is calculated independently. * At least one level is required; multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution across the different PCSV levels, additional shift table on ≤3, >3 to ≤5, > 5 to ≤10, >10 to ≤20, and > 20 category for baseline vs. post baseline may be provided
AST*	>3 and ≤ 5 ULN and baseline ≤ 3 ULN* >5 and ≤ 10 ULN and baseline ≤ 5 ULN >10 and ≤ 20 ULN and baseline ≤ 10 ULN >20 ULN and baseline ≤ 20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Each category is calculated independently. * At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution across the different PCSV levels, additional shift table on ≤3, >3 to ≤5, > 5 to ≤10, >10 to ≤20, and > 20 category for baseline vs. post baseline may be provided
Alkaline Phosphatase	>1.5 ULN and baseline ≤ 1.5 ULN	Enzyme activity must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007.

Parameter	PCSV	Comments
Total Bilirubin*	>1.5 and ≤ 2 ULN and baseline ≤ 1.5 ULN* >2 ULN and baseline ≤ 2.0 ULN	Must be expressed in ULN, not in µmol/L or mg/L. Categories are cumulative. Concept paper on DILI – FDA draft Guidance Oct 2007. * At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution of significant level, additional shift table on ≤1.5, >1.5 to ≤2.0 and > 2.0 category for baseline vs. post baseline may be provided
Conjugated Bilirubin	>35% Total Bilirubin and TBILI>1.5 ULN, and baseline Total Bilirubin ≤ 35% or TBILI ≤1.5 ULN	Conjugated bilirubin determined on a case-by-case basis.
ALT and Total Bilirubin	ALT>3 ULN and TBILI>2 ULN, and baseline ALT ≤ 3 ULN or TBILI ≤ 2ULN	Concept paper on DILI – FDA draft Guidance Oct 2007.
CPK*	>3 and ≤ 10 ULN and baseline ≤ 3ULN* >10 ULN and baseline ≤10ULN	FDA Feb 2005. Am J Cardiol April 2006. Categories are cumulative. * At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution of significant level, additional shift table on ≤3, >3 to ≤10, and > 10 category for baseline vs. post baseline may be provided
Creatinine	≥150 µmol/L (Adults) or ≥ULN (if ULN≥150 µmol/L) and baseline < 150 µmol/L or <ULN (if ULN≥150 µmol/L) ≥30% change from baseline ≥100% change from baseline	Benichou C., 1994. 3 independent criteria

Parameter	PCSV	Comments
Creatinine Clearance (Cockcroft's formula)	<15 ml/min and baseline ≥15 ml/min (end stage renal impairment) ≥15 -<30 ml/min and baseline ≥30 ml/min (severe renal impairment) ≥30 - < 60 ml/min and baseline ≥60 ml/min (moderate renal impairment) ≥60 - < 90 ml/min and baseline ≥90 ml/min (mild renal impairment)	Use is optional. FDA draft guidance 2010 Four independent criteria, will provide additional shift table if needed
Uric Acid Hyperuricemia: Hypouricemia:	 >408 µmol/L or >ULN (if ULN≥408 µmol/L) and baseline ≤408 µmol/L or ≤ULN (if ULN≥408 µmol/L) <120 µmol/L or <LLN (if LLN≤120 µmol/L) and baseline ≥ 120 µmol/L or ≥LLN (if LLN≤120 µmol/L)	Harrison- Principles of Internal Medicine 17th Ed., 2008. Two independent criteria
Blood Urea Nitrogen	≥17 mmol/L or ≥ULN (if ULN≥17 mmol/L) and baseline <17 mmol/L or <ULN (if ULN≥17 mmol/L)	Two independent criteria
Chloride Hypochloremia: Hyperchloremia:	 <80 mmol/L or <LLN (if LLN≤80 mmol/L) and baseline ≥ 80 mmol/L or ≥LLN (if LLN≤80 mmol/L) >115 mmol/L or >ULN (if ULN≥115 mmol/L) and baseline ≤ 115 mmol/L or ≤ULN (if ULN≥115 mmol/L)	Two independent criteria
Sodium Hyponatremia: Hypernatremia:	 ≤129 mmol/L or ≤LLN (if LLN≤129 mmol/L) and baseline > 129 mmol/L or >LLN (if LLN≤129 mmol/L) ≥160 mmol/L or ≥ULN (if ULN≥160 mmol/L) and baseline <160 mmol/L or <ULN (if ULN≥160 mmol/L)	Two independent criteria

Parameter	PCSV	Comments
Potassium Hypokalemia Hyperkalemia	 <3 mmol/L or <LLN (if LLN≤3 mmol/L) and baseline ≥ 3 mmol/L or ≥LLN (if LLN≤3 mmol/L) ≥5.5 mmol/L or ≥ULN (if ULN≥5.5 mmol/L) and baseline <5.5 mmol/L or <ULN (if ULN≥5.5 mmol/L)	FDA Feb 2005. Two independent criteria
Total Cholesterol	≥7.74 mmol/L or ≥ULN (if ULN≥7.74 mmol/L) and baseline < 7.74 mmol/L or <ULN (if ULN≥7.74 mmol/L)	Threshold for therapeutic intervention.
Triglycerides	≥4.6 mmol/L or ≥ULN (if ULN≥4.6 mmol/L) and baseline < 4.6 mmol/L or <ULN (if ULN≥4.6 mmol/L)	Threshold for therapeutic intervention.
Lipasemia	≥3 ULN and baseline < 3 ULN	
Amylasemia	≥3 ULN and baseline < 3 ULN	
Glucose Hypoglycaemia Hyperglycaemia	 ≤3.9 mmol/L and <LLN and baseline >3.9 mmol/L or ≥ LLN ≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted) and baseline < 11.1 mmol/L (unfasted); <7 mmol/L (fasted)	 ADA Jan 2008.
HbA1c	>8% and baseline ≤ 8%	
Albumin	≤25 g/L or ≤LLN (if LLN≤25 g/L) and baseline >25 g/L or >LLN (if LLN≤25 g/L)	
CRP	>2 ULN or >10 mg/L (if ULN not provided) and baseline ≤2 ULN or ≤10 mg/L (if ULN not provided)	FDA Sept 2005.
Hematology		

Parameter	PCSV	Comments
WBC	<3.0 Giga/L or <LLN (if LLN≤3.0 Giga/L) and baseline ≥3.0 Giga/L or ≥LLN (if LLN≤3.0 Giga/L) (Non-Black); <2.0 Giga/L or <LLN (if LLN≤2.0 Giga/L) and baseline ≥2.0 Giga/L or ≥LLN (if LLN≤2.0 Giga/L) (Black)* ≥16.0 Giga/L or ≥ULN (if ULN≥16.0 Giga/L) and baseline < 16 Giga/L or <ULN (if ULN≥16.0 Giga/L)	Increase in WBC: not relevant. *The default criteria. Summary by race (black and Non-black) are optional. To be interpreted only if no differential count available.
Lymphocytes	>4.0 Giga/L or >ULN (if ULN≥4.0 Giga/L) and baseline ≤ 4.0 Giga/L or ≤ULN (if ULN≥4.0 Giga/L)	
Neutrophils	<1.5 Giga/L or <LLN (if LLN≤1.5 Giga/L) for Non-Black or <1.0 Giga/L or <LLN (if LLN≤1.0 Giga/L) for Black and baseline ≥1.5 Giga/L or ≥LLN (if LLN≤1.5 Giga/L) for Non-Black or ≥1.0 Giga/L or ≥LLN (if LLN≤1.0 Giga/L) for Black* <1.5 Giga/L or <LLN (if LLN≤1.5 Giga/L) and baseline ≥1.5 Giga/L or ≥LLN (if LLN≤1.5 Giga/L) (Non-Black); <1.0 Giga/L or <LLN (if LLN≤1.0 Giga/L) and baseline ≥1.0 Giga/L or ≥LLN (if LLN≤1.0 Giga/L) (Black) <500 Giga/L regardless of baseline value or racce	International Consensus meeting on drug-induced blood cytopenias, 1991. *The default criteria. By race (black and Non-black) are optional.
Monocytes	>0.7 Giga/L or >ULN (if ULN≥0.7 Giga/L) and baseline ≤ 0.7 Giga/L or ≤ULN (if ULN≥0.7 Giga/L)	
Basophils	>0.1 Giga/L or >ULN (if ULN≥0.1 Giga/L) and baseline ≤ 0.1 Giga/L or ≤ULN (if ULN≥0.1 Giga/L)	
Eosinophils	>0.5 Giga/L or >ULN (if ULN≥0.5 Giga/L) and baseline ≤0.5 Giga/L or ≤ULN (if ULN≥0.5 Giga/L)	Harrison- Principles of Internal Medicine 17th Ed., 2008.

Parameter	PCSV	Comments
Hemoglobin	≤115 g/L or ≤LLN (if LLN≤115 g/L) for male or ≤95 g/L or ≤LLN (if LLN≤95 g/L) for female and baseline > 115 g/L or >LLN (if LLN≤115 g/L) for male or > 95 g/L or >LLN (if LLN≤95 g/L) for Female* ≤115 g/L or ≤LLN (if LLN≤115 g/L) and baseline > 115 g/L or >LLN (if LLN≤115 g/L) for male; ≤95 g/L or ≤LLN (if LLN≤95 g/L) and baseline > 95 g/L or >LLN (if LLN≤95 g/L) for Female. ≥185 g/L or ≥ULN (if ULN≥185 g/L) for male or ≥165 g/L or ≥ULN (if ULN≥165 g/L) for female and baseline <185 g/L or <ULN (if ULN≥185 g/L) for male or <165 g/L or <ULN (if ULN≥165 g/L) for Female* ≥185 g/L or ≥ULN (if ULN≥185 g/L) and baseline <185 g/L or <ULN (if ULN≥185 g/L) for Male; ≥165 g/L or ≥ULN (if ULN≥165 g/L) and baseline < 165 g/L or <ULN (if ULN≥165 g/L) for Female Decrease from Baseline ≥20 g/L	Three criteria are independent. *The default criteria. By gender (male and female) are optional. Criteria based upon decrease from baseline are more relevant than based on absolute value. Other categories for decrease from baseline can be used (≥30 g/L, ≥40 g/L, ≥50 g/L).

Parameter	PCSV	Comments
Hematocrit	≤0.37 v/v or ≤LLN (if LLN≤0.37 v/v) for Male or ≤0.32 v/v or ≤LLN (if LLN≤0.32 v/v) for Female and baseline > 0.37 v/v or >LLN (if LLN≤0.37 v/v) for Male or > 0.32 v/v or >LLN (if LLN≤0.32 v/v) for Female* ≤0.37 v/v or ≤LLN (if LLN≤0.37 v/v) and baseline > 0.37 v/v or >LLN (if LLN≤0.37 v/v) for Male ; ≤0.32 v/v or ≤LLN (if LLN≤0.32 v/v) and baseline > 0.32 v/v or >LLN (if LLN≤0.32v/v) for Female ≥0.55 v/v or ≥ULN (if ULN≥0.55 v/v) for Male or ≥0.5 v/v or ≥ULN (if ULN≥0.5 v/v) for Female and baseline < 0.55 v/v or <ULN (if ULN≥0.55 v/v) for Male < 0.5 v/v or <ULN (if ULN≥0.5 v/v) for Female* ≥0.55 v/v or ≥ULN (if ULN≥0.55 v/v) and baseline < 0.55 v/v or <ULN (if ULN≥0.55 v/v) for Male ; ≥0.5 v/v or ≥ULN (if ULN≥0.5 v/v) and baseline < 0.5 v/v or <ULN (if ULN≥0.5 v/v) for Female	Two Criteria are independent *The default criteria. By gender (male and female) are optional.
RBC	≥6 Tera/L or ≥ULN (if ULN≥6 Tera/L) and baseline < 6 Tera/L or <ULN (if ULN≥6 Tera/L)	Unless specifically required for particular drug development, the analysis is redundant with that of Hb.
Platelets	<100 Giga/L or <LLN (if LLN≤100 Giga/L) and baseline ≥100 Giga/L or ≥LLN (if LLN≤100 Giga/L) ≥700 Giga/L or ≥ULN (if ULN≥700 Giga/L) and baseline < 700 Giga/L or <ULN (if ULN≥700 Giga/L)	International Consensus meeting on drug-induced blood cytopenias, 1991. Two independent criteria
Urinalysis		

Parameter	PCSV	Comments
pH	≤4.6 or ≤LLN (if LLN≤4.6) and baseline > 4.6 or >LLN (if LLN≤4.6) ≥8 or ≥ULN (if ULN≥8) and baseline < 8 or <ULN (if ULN≥8)	Two independent criteria
Vital signs		
HR	<45 bpm and decrease from baseline ≥20 bpm ≥120 bpm and increase from baseline≥20 bpm	To be applied for all positions except STANDING
SBP	≤95 mmHg and decrease from baseline ≥20mmHg ≥160 mmHg and increase from baseline ≥20 mmHg	To be applied for all positions except STANDING
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg ≥110 mmHg and increase from baseline ≥10 mmHg	To be applied for all positions except STANDING
Weight	≥ 5% increase from baseline ≥5% decrease from baseline	FDA Feb 2007

10.3. List of Antipyretics

To search antipyretics in the datasets, ATC codes 'M01A' and 'N02B' and below standardized medication names, as well as 'PARACETAMOL' and 'acetaminophen' may be used.

STANDARDIZED MEDICATION NAME:

ACECLOFENAC

ACEMETACIN

ACETYLSALICYLATE LYSINE

ACETYLSALICYLIC ACID

ALOE VERA;METHYL SALICYLATE

AMTOLMETIN GUACIL

BENZYDAMINE

BENZYDAMINE HYDROCHLORIDE

BENZYDAMINE HYDROCHLORIDE;CHLORHEXIDINE GLUCONATE

CAFFEINE;CODEINE PHOSPHATE;METAMIZOLE SODIUM;NAPROXEN;PHENOBARBITAL

CAFFEINE;DROTAVERINE HYDROCHLORIDE;NAPROXEN;PARACETAMOL;PHENIRAMINE MALEATE

CELECOXIB

CHLORPHENAMINE MALEATE;IBUPROFEN;PSEUDOEPHEDRINE HYDROCHLORIDE

CHLORZOXAZONE;DICLOFENAC SODIUM;PARACETAMOL

CHOLINE SALICYLATE

CODEINE PHOSPHATE HEMIHYDRATE;PARACETAMOL

CODEINE PHOSPHATE;DICLOFENAC SODIUM

CODEINE PHOSPHATE;IBUPROFEN

CODEINE PHOSPHATE;IBUPROFEN;PARACETAMOL

CODEINE;IBUPROFEN;PARACETAMOL

CYANOCOBALAMIN;DICLOFENAC SODIUM;PYRIDOXINE HYDROCHLORIDE;THIAMINE HYDROCHLORIDE

CYCLOBENZAPRINE HYDROCHLORIDE;LIDOCAINE;PIROXICAM

DEXAMETHASONE;PHENYLBUTAZONE

DEXIBUPROFEN

DEXKETOPROFEN

DEXKETOPROFEN TROMETAMOL

DIACEREIN

DICLOFENAC

DICLOFENAC DIETHYLAMINE

DICLOFENAC POTASSIUM

DICLOFENAC POTASSIUM;PARACETAMOL;SERRAPEPTASE

DICLOFENAC SODIUM

DICLOFENAC SODIUM;HEPARIN SODIUM

DICLOFENAC SODIUM;LIDOCAINE HYDROCHLORIDE

DICLOFENAC SODIUM;LINUM USITATISSIMUM;MENTHOL;METHYL SALICYLATE;THIOCOLCHICOSIDE

DICLOFENAC SODIUM;MISOPROSTOL

DICLOFENAC SODIUM;OMEPRAZOLE

DICLOFENAC SODIUM;PARACETAMOL

DICLOFENAC;MISOPROSTOL

DICYCLOVERINE;PARACETAMOL

DIFLUNISAL

DIMETHYL SULFOXIDE

DIPHENHYDRAMINE HYDROCHLORIDE;NAPROXEN SODIUM

DIPHENHYDRAMINE;IBUPROFEN

ESOMEPRAZOLE MAGNESIUM;NAPROXEN

ETODOLAC

ETOFENAMATE

ETORICOXIB

FAMOTIDINE;IBUPROFEN

FENOPROFEN

FENOPROFEN CALCIUM

FLURBIPROFEN

HYDROCODONE;IBUPROFEN

IBUPROFEN

IBUPROFEN;LEVOMENTHOL

IBUPROFEN;METHOCARBAMOL

IBUPROFEN;PARACETAMOL

IBUPROFEN;PSEUDOEPHEDRINE

IBUPROFEN;PSEUDOEPHEDRINE HYDROCHLORIDE

INDOMETACIN

KETOPROFEN

KETOROLAC

KETOROLAC TROMETHAMINE

LORNOXICAM

MEFENAMIC ACID

MELOXICAM

METHYL SALICYLATE

NABUMETONE

NAPROXEN

NAPROXEN SODIUM

NAPROXEN SODIUM;PSEUDOEPHEDRINE HYDROCHLORIDE

NEPAFENAC

NIFLUMIC ACID

NIMESULIDE

OTHER ANTIINFLAMMATORY AND ANTIRHEUMATIC AGENTS, NON-STEROIDS

OXAPROZIN

PARECOXIB

PARECOXIB SODIUM

PHENYLBUTAZONE

PIROXICAM

PIROXICAM BETADEX

SALICYLATE SODIUM

SALICYLIC ACID

SALSALATE

SALSALATE

SULINDAC

TENOXICAM

VIMOVO

10.4. Conversion from Oxygen flow rate to estimated FiO_2 (fraction inspired oxygen)

Method (Oxygen delivery device)	Oxygen flow rate O_2 flow (liters/min)	Fraction of Inspired Oxygen [FiO_2] (%)
Nasal cannula	1	24
	2	28
	3	32
	4	36
	5	40
	6	44
	>=7	Impute as (O_2*4)+20. Max is 100
Nasopharyngeal catheter	4	40
	5	50
	6	60
	>=7	Impute as O_2*10. Max is 100
Face mask	5	40
	6	50
	7	55 (impute)
	8	60
	9	65 (impute)
	>=10	70 (impute)
Face mask with reservoir (also called non-breather face mask)	<6	60 (impute)
	6	60
	7	70
	8	80
	9	90
	10	95
	>=10	95 (impute)
OTHER	Any value	Keep as recorded or NULL (no imputation)

Note: If FiO_2 data is entered in the CRF for above devices, ignore value and derive as above. If no supplemental oxygen is used, then derive FiO_2 as 21%. For Other devices, data query may be done. Nasopharyngeal catheter is not a pre-specified option on the CRF page.

Signature Page for VV-RIM-00103288 v1.0

ESig Approval	[redacted] 13-Apr-2020 16:07:33 GMT+0000
ESig Approval	[redacted] 13-Apr-2020 16:12:48 GMT+0000
ESig Approval	[redacted]ment 13-Apr-2020 16:12:59 GMT+0000
ESig Approval	[redacted]ment 13-Apr-2020 16:31:20 GMT+0000
ESig Approval	[redacted] 13-Apr-2020 16:56:11 GMT+0000
ESig Approval	[redacted]ment 13-Apr-2020 17:04:11 GMT+0000

Signature Page for VV-RIM-00103288 v1.0 Approved

REGENERON
Biostatistics and Data Management

STATISTICAL ANALYSIS PLAN
FOR PHASE 3 PORTION OF THE STUDY
VERSION: FINAL

AN ADAPTIVE PHASE 2/3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY ASSESSING EFFICACY AND SAFETY OF SARILUMAB FOR HOSPITALIZED PATIENTS WITH COVID-19

Compound:	REGN88 (sarilumab, Kevzara®)
Protocol Number:	6R88-COV-2040
Clinical Phase:	Phase 2/3
Sponsor:	Regeneron Pharmaceuticals, Inc.
Study Biostatisticians:	[redacted]
Clinical Trial Manager:	[redacted]
Study Medical Directors:	[redacted]
Version/Date:	Original (Version 1.0) / 16 MAY 2020

The approval signatures below indicate that these individuals have reviewed the Statistical Analysis Plan (SAP) and agreed on the planned analysis defined in this document for reporting.

See appended electronic signature page

Study Biostatisticians [redacted]

See appended electronic signature page

Study Pharmacokinetics [redacted]

See appended electronic signature page

Study Medical Directors [redacted]

See appended electronic signature page

Head of BDM or designee [redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Definition
AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
ANC	Absolute Neutrophil Count
ANCOVA	Analysis of covariance
ARDS	Acute respiratory distress syndrome
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
COVID-19	Coronavirus Disease 2019
CPK	Creatine phosphokinase
CRF	Case report form (electronic or paper)
CRP	C-reactive protein
ECG	Electrocardiogram
ECMO	Extracorporeal membrane oxygenation
FiO_2	Fraction inspired oxygen
ICH	International Council for Harmonisation
ICU	Intensive care unit
IDMC	Independent Data Monitoring Committee
IFN	Interferon
IL-6	Interleukin 6
IRT	Interactive Response Technology
ITT	Intention-to-treat
IWRS	Interactive Web Response System (web system of IRT)
IV	Intravenous
mAb	Monoclonal Antibody
mITT	Modified intention-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
MERS-CoV	Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV)

Abbreviation	Definition
NCI-CTCAE	National Cancer Institute-Common Terminology Criteria for Adverse Events
OP	Oropharyngeal
NP	Nasopharyngeal
PA5	Protocol Amendment #6 incorporating Phase 3 adaptations
PCSV	Potentially Clinically Significant Value
PD	Pharmacodynamic(s)
PK	Pharmacokinetic(s)
PPS	Per Protocol population set
PT	Preferred Term
RA	Rheumatoid Arthritis
RBC	Red Blood Cell
RNA	Ribonucleic Acid
Regeneron	Regeneron Pharmaceuticals, Inc.
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SAS	Statistical Analysis System
SC	Subcutaneous
SOC	System organ class
SpO_2	Peripheral capillary oxygen saturation
SUSAR	Suspected unexpected serious adverse reaction
TEAE	Treatment-emergent adverse event
ULN	Upper Limit Normal
US	United States (of America)
VFD	Ventilator-free days
WBC	White blood cell
WHO	World Health Organization
WHODD	World Health Organization Drug Dictionary

1. OVERVIEW

The purpose of the statistical analysis plan (SAP) is to ensure the integrity of the study results by pre-specifying the statistical approaches for the analysis of study data prior to the Phase 3 database lock of the Phase 2/3 adaptive study 6R88-COV-2040 of sarilumab (Kevzara®) in patients with COVID-19. This Phase 3 SAP is intended to be a comprehensive and detailed description of the strategy and statistical methods to be used in the analysis of data for the Phase 3 portion of the study based on Protocol Amendment 6 (PA6) (dated 16-MAY-2020).

The SAP for the Phase 2 portion of the study (called Phase 2 SAP) was finalized on 13-APR-2020 prior to the first interim database lock for Phase 2 on 15-APR-2020. The last patient in Phase 2 was randomized and dosed on 31-MAR-2020. Phase 2 included ~460 patients randomized (2:2:1) to sarilumab 400 mg IV, sarilumab 200 mg IV or placebo in all COVID-19 severity strata: severe, critical, and multi-organ system dysfunction (MSOD) (including immunocompromised).

Beginning on 01-APR-2020 patients began enrolling into the Phase 3 portion of the study into the same treatment arms and disease severity strata as Phase 2 until 27-APR-2020 when changes were made based on recommendations from the Independent Data Monitoring Committee. These changes and additional adaptations were incorporated into Protocol Amendment 6 (16-May-2020).

The Phase 2 and Phase 3 portions of this study will be analyzed separately.

The main focus of this Phase 3 SAP is analysis of the following sets of patients described as cohorts, which were comprised of subjects of differing severity, randomized to differing treatment regimens:

1. Cohort 1: Consists of all Phase 3 patients randomized to sarilumab 400 mg IV, sarilumab 200 mg IV or placebo across disease severity strata (severe, critical, multi-system organ dysfunction or immunocompromised). Within this Cohort, the primary focus of this SAP is on the efficacy and safety data on patients in the critical stratum randomized to sarilumab 400 mg or placebo. For patients in the critical stratum randomized to sarilumab 200mg, only safety data will be analyzed.
2. Cohort 2: Cohort of hospitalized COVID-19 patients to be randomized to sarilumab 800 mg or placebo who are receiving mechanical ventilation at baseline.
3. Cohort 3: Cohort of hospitalized COVID-19 patients to be randomized to sarilumab 800 mg or placebo who are receiving high-intensity oxygen therapy without mechanical ventilation at baseline.

Analyses of patients previously enrolled in Phase 3Cohort 1 randomized to sarilumab 400 mg IV, sarilumab 200 mg IV or placebo will be done for each disease severity stratum of severe, MSOD, or immunocompromised, for exploratory purpose only. These exploratory analyses will be designed similarly as the main analyses described in this SAP. Exploratory analyses will be summarized in one or more clinical study reports.

1.1. Background/Rationale

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel enveloped RNA betacoronavirus that emerged in December 2019 in Wuhan, China. The term COVID-19 is the disease caused by SARS-CoV-2 with symptoms that manifest a median of 5 days and up to 14 days after infection. In March 2020, the World Health Organization officially declared COVID-19 a pandemic. (Hereafter, SARS-CoV2 and COVID-19 are interchangeably used.) The most frequent clinical presentation of severe COVID-19 is pneumonia with symptoms including fever, cough, and dyspnea (shortness of breath). Acute viral pneumonia is a major cause of morbidity in patients with COVID-19 and patients can be at risk of developing acute respiratory distress syndrome (ARDS), a syndrome of severe impairment of gas exchange. Clinically, ARDS presents with severe hypoxemia (abnormally low levels of oxygen in the blood).

A key driver of this deterioration is an overactive inflammatory response in the lungs (Mehta, 2020 [10]). Seriously ill patients infected by SAR-CoV-2 were found to have elevated circulatory inflammatory mediators such as C-reactive protein (CRP), IL-1B, IL-1RA, IL-7, IL-8, IL-9, IFNγ, IFN-γ-inducible protein (IP10), monocyte chemoattractant protein (MCP1), tumor necrosis factor (TNFα) (Huang, 2020 [5]), resembling the cytokine storm observed in similar lower respiratory disease caused by infections due to SARS-CoV, identified in 2002, and by the Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV) (Zhang, 2020 [14]). COVID-19 has been associated with pulmonary release of cytokines such as interleukin-6 (IL-6) which has been shown to be a major contributor to the development of fever and hypoxemia. One hypothesis is that IL-6 is a modifiable driver of progression of COVID-19 pneumonia.

Therefore, a potential therapeutic approach would be to attenuate the overactive inflammatory immune response in the lungs through blockade of the IL-6 receptor. In the current study, sarilumab (Kevzara®), an anti-IL-6 receptor (anti-IL-6R) monoclonal antibody (mAb), is being evaluated for efficacy and safety in treatment of patients with COVID-19 at the doses of 200 mg intravenous (IV), 400 mg IV and 800 mg IV. Kevzara® (sarilumab) has been previously approved for the treatment of rheumatoid arthritis (RA) at 200 mg Q2W (subcutaneous [SC]) [with down dosing to 150 mg Q2W (SC) for certain laboratory changes]. Sarilumab is highly similar to tocilizumab (Actemra®), another anti-IL-6R mAb, and there is reported evidence from a preliminary uncontrolled study in China in 21 patients that IV treatment with 400 mg of tocilizumab provided a clinically meaningful improvement in clinical symptoms that are thought to be mediated by cytokine release in patients with severe or critical COVID-19 infection (Xu, 2020 [12]).

At the time of the start of this study, there were no specific COVID-19 treatments. Hospitalized patients were managed clinically through supportive therapy including supplemental oxygen therapy, empiric treatments and intensive care as required. Many therapeutic agents have been used to treat patients with SARS-CoV and MERS-CoV, including corticosteroids, type 1 interferon (IFN) agents, convalescent plasma, ribavirin, lopinavir/ritonavir, proteases, and agents targeting viral entry proteins. However, none have been proven to be efficacious in clinical trials at the time of the initiation of this study, justifying the use of placebo as a comparator for this study.

This study 6R88-COV-2040 is designed as a randomized well-controlled Phase 2/3 clinical trial in hospitalized patients across multiple disease severity stages of COVID-19. Due to the novel nature of COVID-19, efficacy endpoints are not well established. The Phase 2/3 adaptive design of the study allows for identification of early signs of clinical efficacy in the Phase 2 data (e.g., reduction in C-Reactive Protein, estimates of effects on clinical status), which will allow adaptation of the Phase 3 study design including: 1) selection of clinical outcome endpoints in Phase 3 (e.g., well defined endpoints based on clinical status assessment of patients), 2) confirming the Phase 3 population, 3) evaluate further doses of sarilumab and 4) sample size adjustment for Phase 3.

In the Phase 2 portion of the study, patients were randomized in a 2:2:1 allocation ratio to sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo in 3 disease severity strata: severe, critical and multi-system organ dysfunction (MSOD). The Phase 3 portion of the study seamlessly enrolled patients similarly as in Phase 2 until a Phase 2 interim analysis was reviewed, which informed adaptations to the Phase 3 design.

Following review of the Phase 2 interim data, and based on IDMC recommendations, enrollment in Phase 3 continued in only the 400 mg sarilumab dose and placebo arms and only in the critical stratum. Adaptations to the study design also included division of Phase 3 into 3 cohorts as previously described in Section 1. Enrollment will continue in Cohort 1 until approximately 170 patients randomized to 400 mg or placebo in the critical stratum who were on mechanical ventilation at baseline have been treated. Cohort 2 which consists of patients who are on mechanical ventilation at baseline and Cohort 3 which consists of patients not on mechanical ventilation but on high intensity oxygen therapy at baseline, will not begin enrollment until after Cohort 1 enrollment has completed. The new Cohorts 2 and 3 will be randomized 1:1 to sarilumab 800 mg or placebo. Further details of the Phase 2/3 adaptive study design are described in Section 2.1.

1.2. Study Objectives for Phase 3

1.2.1. Primary Objectives

Phase 3 Cohort 1:

The primary objective of the study is to evaluate the clinical efficacy of sarilumab relative to the control arm in adult patients hospitalized with critical COVID-19 on mechanical ventilation at baseline.

Phase 3 Cohort 2:

The primary objective of the study is to evaluate the clinical efficacy of sarilumab relative to the control arm in adult patients hospitalized with COVID-19 receiving mechanical ventilation at baseline.

Phase 3 Cohort 3:

The primary objective of the study is to evaluate the clinical efficacy of sarilumab relative to the control arm in adult patients hospitalized with COVID-19 receiving high-intensity oxygen therapy* without mechanical ventilation at baseline.

* High intensity oxygen therapy is defined as the use of non-rebreather mask with an oxygen flow rate of at least 10 L/min; use of a high flow device with at least 50% FiO2, or use of non-invasive ventilation (eg, BiPAP™) or continuous positive airway pressure (CPAP) to treat hypoxemia.

1.2.2. Secondary Objectives

Phase 3 Efficacy

The secondary efficacy objectives of the study are to:

Phase 3 Cohort 1:

- Determine whether sarilumab improves respiratory outcomes in patients with critical COVID-19
- Determine whether sarilumab reduces mortality in patients with critical COVID-19
- Determine whether sarilumab shortens hospitalization in patients with critical COVID-19

Phase 3 Cohort 2:

- Determine whether sarilumab improves respiratory outcomes in patients with COVID-19 receiving mechanical ventilation
- Determine whether sarilumab reduces mortality in patients with COVID-19 receiving mechanical ventilation
- Determine whether sarilumab shortens hospitalization in patients with COVID-19 receiving mechanical ventilation

Phase 3 Cohort 3:

- Determine whether sarilumab improves respiratory outcomes in patients with COVID-19 receiving high intensity oxygen therapy without mechanical ventilation
- Determine whether sarilumab reduces mortality in patients with COVID-19 receiving high intensity oxygen therapy without mechanical ventilation
- Determine whether sarilumab shortens hospitalization in patients with COVID-19 receiving high intensity oxygen therapy without mechanical ventilation

Phase 3 Safety

Safety

The secondary safety objectives of the study are to evaluate the safety of sarilumab through hospitalization (up to Day 29 if patient is still hospitalized) compared to the control arm as assessed by incidence of:

- Serious adverse events (SAEs)
- Grade 4 neutropenia (absolute neutrophil count [ANC]<500/mm3)

- Grade 4 neutropenia (ANC<500/mm3) with concurrent severe or life-threatening bacterial, invasive fungal, or opportunistic infection
- Grade ≥2 infusion related reactions
- Grade ≥2 hypersensitivity reactions
- Increase in alanine transaminase (ALT) or aspartate aminotransferase (AST) ≥3X upper limit of normal (ULN) (for patients with normal baseline) or >3X ULN AND at least 2 fold increase from baseline value (for patients with abnormal baseline)
- Invasive bacterial or fungal infections of clinical significance with confirmed diagnosis based on the investigator's assessment with appropriate diagnostic workups and consultations

1.2.3. Modifications from the Statistical Section in the Final Protocol

Not applicable.

1.2.4. Revision History for SAP Amendments

Not applicable.

2. INVESTIGATION PLAN

2.1. Study Design and Randomization

This study is an adaptive Phase 2/3, randomized, double-blind, placebo-controlled trial to evaluate the efficacy and safety of sarilumab compared with placebo in hospitalized adults with severe or critical COVID-19 (as broadly defined below).

At the time of the original Phase 3 SAP, the study is ongoing and being conducted in the United States (US) in up to 150 sites. Enrollment in the study began on 17 Mar 2020.

Phase 2 Study Design

In the Phase 2 portion of the study, a total of approximately 460 patients were randomized in a 2:2:1 allocation ratio to sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo in a stratified manner up to 26 Apr 2020. A Phase 2 interim analysis was conducted on 27 Apr 2020 whereby upon recommendation of the Independent Data Monitoring Committee (IDMC), adaptations were made to the Phase 3 portion of the study (described below).

In Phase 2 randomization was stratified by

- Severity of illness at enrollment
  - Severe disease
  - Critical disease
  - Multi-system organ dysfunction
    - Immunocompromised†
- Systemic corticosteroids (Yes/No)

† Immunocompromised patients were permitted to be enrolled in this study after Protocol Amendment #3 (28-MAR-2020) at the request of the FDA. For Phase 2 analysis, these patients will be reported as a subset of the Multi-system organ dysfunction strata.

The severity categories are:

1. **Severe disease**
   - Requires supplemental oxygen administration by nasal cannula, simple face mask, or other similar oxygen delivery device
2. **Critical disease**
   - Requires supplemental oxygen requiring delivered by non-rebreather mask or high-flow nasal cannula, OR
   - Use of invasive or non-invasive ventilation, OR
   - Requiring treatment in an intensive care unit.
3. **Multi-system organ dysfunction**
   - Multi-system organ dysfunction: use of vasopressors, extracorporeal life support, or renal replacement therapy

4. **Immunocompromised**
    - Immunocompromised patients (or on immunosuppressant treatments)

Phase 3 Study Design

For Phase 3 Cohort 1, randomization was similarly stratified as in Phase 2, by disease severity at enrollment (severe, critical, MSOD, immunocompromised) and use of systemic corticosteroids (yes/no)

As part of the Phase 3 adaptation plan, the following adaptations were made to the Phase 3 study design (see Table 1). Refer to Protocol Section 3.2.2 for the interim Phase 2 results and rationale for study design adaptation.

Table 1: Summary of Phase 3 Adaptations

Study Component	Adaptation
Phase 3 endpoint	At least 1-point improvement from baseline in clinical status on the 7-point ordinal scale (For example, a 1-point improvement from baseline in clinical status for a patient in Cohort 1 in critical stratum who is on mechanical ventilation at baseline, will imply that the patient is now off mechanical ventilation. This kind of a treatment effect measure by a 1-point improvement on the ordinal scale is considered clinically important.)
Phase 3 primary analysis population	• Cohort 1: Critical patients receiving mechanical ventilation at baseline • Cohort 2: Patients on mechanical ventilation at baseline • Cohort 3: Patients on high intensity oxygen therapy without mechanical ventilation at baseline*
Dose regimens	• Cohort 1: Primary comparison will be between sarilumab 400 mg and placebo • Cohort 2: Sarilumab 800 mg and placebo • Cohort 3: Sarilumab 800 mg and placebo
Sample size for the Primary Analysis Population in Cohort 1	~170
Sample Size for Cohort 2	225
Sample Size for Cohort 3	225

* High intensity oxygen therapy is defined as the use of non-rebreather mask with an oxygen flow rate of at least 10 L/min; use of a high flow device with at least 50% FiO2, or use of non-invasive ventilation (eg, BiPAP™) or CPAP to treat hypoxemia.

Phase 3 Cohort 1

Until 26 Apr 2020, the Phase 3 portion of the study randomized patients in a stratified manner similar to the Phase 2 study design described above. As of 27 Apr 2020 (as per recommendation of the IDMC), patients in Phase 3 Cohort 1 were randomized to sarilumab 400 mg or placebo

only in the critical stratum in a 2:1 manner. Based on the IDMC recommendation, patients in the severe and MSOD strata were no longer enrolled nor received study drug.

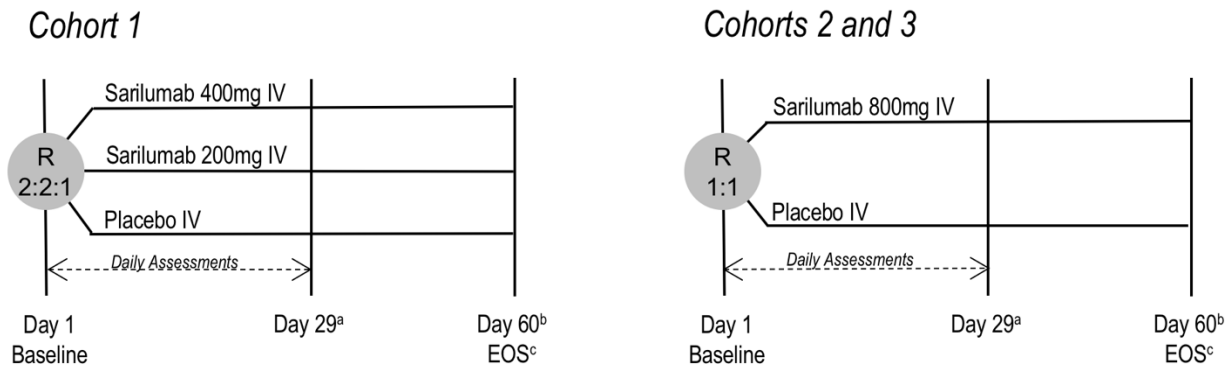
Enrollment into Cohort 1 will end approximately 22 days after the 170th patient in the critical stratum on mechanical ventilation at baseline who were randomized to receive either sarilumab 400 mg or placebo is enrolled.

Phase 3 Cohort 2 and Cohort 3

Cohort 2 and Cohort 3 will enroll in parallel when Cohort 1 enrollment is completed. In Cohorts 2 and 3, patients will be randomized 1:1 to receive sarilumab 800 mg IV or placebo stratified by:

- Use of a non-IL6/6R therapy administered under an Emergency Use Authorization (EUA) at randomization (yes/no)
- Use of systemic corticosteroids at randomization (yes/no)

Figure 1: Phase 3 Study Design – Cohorts 1, 2 and 3



2.2. Sample Size and Power Considerations

For the Phase 2 portion of the study, a total of approximately 460 patients were randomized (2:2:1) to sarilumab 400 mg IV, sarilumab 200 mg IV or placebo. This sample size was estimated to provide at least 90% power to detect a treatment effect in the Phase 2 endpoint of reduction in CRP levels as well as allow adequate estimation and improve precision for selection of Phase 3 efficacy endpoints.

The data lock point for the first interim analysis of the Phase 2 portion of the study was set at 15 days after the last of the approximately 460 patients in Phase 2 had been randomized. These data were used to determine the adaptations and analysis plan for the Phase 3 portion of this adaptive Phase 2/3 study. Based on the Phase 2 interim results, the sample size for Cohort 1 of Phase 3 was recalculated using the chosen Phase 3 endpoint of the proportion of patients with at least 1 point improvement in clinical status on the ordinal scale at Day 22 in patients in the critical stratum who were on mechanical ventilation at baseline.

In Phase 3 Cohort 1, the hypothesis will be tested only in patients in the critical stratum, first in the subset of patients on mechanical ventilation without ECMO at baseline as an enrichment strategy and then in the overall critical stratum (ITT). In the Phase 2 portion of the study, in a subgroup analysis of the critical stratum based on being on a mechanical ventilator at baseline, in the mechanical ventilator group, approximately 17% of patients treated with placebo (ie, standard of care) and 57% of patients treated with sarilumab 400 mg (on top of standard of care) achieved at least 1-point improvement by Day 22.

Cohort 1:

With a 2:1 randomization ratio (sarilumab 400 mg IV: placebo), an effect size the same as that observed in Phase 2, and 170 patients in the critical stratum on mechanical ventilation, then the comparison between sarilumab 400 mg (n~113) and placebo (n~57) would have >99% power. Assuming the rate on placebo is 17% and the rate on sarilumab 400 mg is 37% (ie, a difference in proportions is one-half that observed in Phase 2), then the sample size of 170 would provide approximately 80% power to detect this reduced difference. These calculations assume α=0.045, allowing for an interim analysis at the 0.005 level. Therefore, this study plans to enroll approximately 450 critical patients in order to have approximately 170 patients within the critical stratum who are on mechanical ventilation and randomized to 400 mg or placebo.

Note that the total number of patients enrolled in Phase 3 Cohort 1 is larger than that to account for the patients who had been randomized to the severe stratum, MSOD stratum, or to the 200 mg dose group before the IDMC decision to discontinue further enrollment into those strata and dose groups.

Cohort 2:

With a 1:1 randomization ratio (sarilumab 800 mg IV: placebo), an effect size the same as that observed in Phase 2 for 400 mg, and approximately 225 patients total, then the comparison between sarilumab 800 mg and placebo (n=~112 for each group) would have >99% power to detect the difference observed in Phase 2, and would have 92% power if the effect were half of that observed in Phase 2 (ie, 20% difference). These calculations assume α=0.045 (2-sided), allowing for an interim analysis at the 0.005 level. A sample size re-estimation may occur after Phase 3 Cohort 1 results are determined.

Cohort 3:

A sample size of 225 subjects randomized 1:1 (sarilumab 800 mg IV:placebo) would have 92% power to detect a difference in proportions of 20%. These calculations assume α=0.045 (2-sided), allowing for an interim analysis at the 0.005 level. A sample size re-estimation may occur after Phase 3 Cohort 1 results are determined.

2.3. Study Plan

Study flow for both Phases 2 and 3 is shown in Figure 2.

Figure 2: Study Flow Diagram for Phase 2 and Phase 3

[a] The EOS will be on day 60 or day of death, whichever comes first.

[b] If the patient has been discharged from the hospital before day 29, the study site staff will contact the patient for a follow-up phone call.

[c] If the patient has been discharged from the hospital before day 60, the study site staff will contact the patient for a follow-up phone call.

EOS: End of study

Patients will be assessed daily while hospitalized and undergo a series of efficacy, safety, and laboratory assessments while in the hospital. Discharged patients will be contacted by telephone on Day 29 to assess status and occurrence of re-admission to a hospital. All patients will have an end of study (EOS) assessment at Day 60 to collect data on survival status and history of hospital re-admission, and this assessment may be done by phone.

The Study event table is presented in Section 10.

Study patients, the principal investigators, and study site personnel continue to remain and will remain blinded to all randomization assignments throughout the Phase 3 portion of study that is covering: 1) Cohort 1 Critical stratum randomized to sarilumab 400 mg or placebo, and 2) Cohorts 2 and 3 randomized to sarilumab 800 mg or placebo. Similarly, for this portion of the Phase 3 study, the Regeneron Medical/Study Director, Study Monitor, study team directly involved with the conduct of the study and any other Regeneron and contract research organization (CRO) personnel who are in regular contact with the study site will remain blinded to patient randomization assignments.

3. ANALYSIS POPULATIONS

In accordance with guidance from the International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline ICH E9 Statistical Principles for Clinical Trials (ICH, 1998 [4]), the following population of analysis will be used for all statistical analysis:

3.1. Intention-to-Treat (ITT) Population

The intention-to-treat (ITT) population includes all randomized patients who received at least one dose of the study drug. Analysis of the ITT population will be done according to the initial treatment assigned to the patient (as randomized).

ITT population will be the primary analysis population for the Phase 3 Cohort 1, Cohort 2 and Cohort 3 and each cohort will be analyzed separately. For Phase 3 Cohort 1, the specific primary population will be the subset of patients in the critical stratum who are on mechanical ventilation without ECMO at baseline and randomized to sarilumab 400 mg or placebo. In addition, for Cohort 1, the ITT population of patients in the critical stratum will be analyzed for some of the key secondary efficacy outcomes. Demographics, baseline characteristics, and patient disposition will be summarized in the ITT population.

3.2. Per Protocol Set (PPS)

The per protocol population set (PPS) includes all ITT patients who do not have any relevant major protocol deviations, ie, patients with relevant major protocol deviations will be excluded from PPS. A relevant major protocol deviation is one that may affect the interpretation of study efficacy results. The final determination of the definition of the PPS will be made prior to the first database lock. Criteria for relevant major protocol deviations are given in Appendix 10.3.

Analysis of the PPS will be done according to the treatment the patient actually received (as treated). Determination of “as treated” will be based on the actual study drug received on Day 1. The PPS will be used only in Phase 3 Cohort 1, Cohort 2 and Cohort 3 for sensitivity analysis of the primary efficacy endpoints.

3.3. Safety Population (SAF)

The Safety population (SAF) includes all randomized patients who received at least one dose of the study drug. Analysis of the Safety population will be done according to the treatment received (as treated). Determination of “as treated” will be based on the actual study drug received on Day 1. The SAF will be used for analysis of all safety data, treatment exposure, medical history and medications use in the Phase 3 Cohort 1 Critical stratum of patients randomized to sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo, and in Phase 3 Cohort 2 and Cohort 3 of patients randomized to sarilumab 800 mg IV or placebo.

3.4. Exploratory ITT population (EITT)

An Exploratory ITT Population will also be defined similarly as the ITT to perform efficacy analyses in the Phase 3 Cohort 1 population of patients in the strata that were discontinued after

Phase 2 interim data press release, namely, severe, MSOD and immunocompromised patients in all 3 treatment groups (sarilumab 400 mg, sarilumab 200 mg or placebo). Analyses will be done separately for each stratum.

3.5. Exploratory Safety Population (ESAF)

An Exploratory Safety Population will also be defined similarly as the SAF to perform safety analyses in the Phase 3 Cohort 1 population of patients in the strata that were discontinued after Phase 2 interim data press release, namely, severe, MSOD, and immunocompromised patients in all 3 treatment groups (sarilumab 400 mg, sarilumab 200 mg, or placebo). Analyses will be done separately for each stratum.

3.6. Pharmacokinetics Analysis Set

The pharmacokinetics (PK) analysis population includes all patients who received any study drug and who had at least 1 non-missing result following the first dose of study drug.

4. ANALYSIS VARIABLES

4.1. Demographic and Baseline Characteristics

Demographic and baseline characteristic variables include the following:

- Age at screening (years)
- Age group (years) (18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, >=85) (lower and upper limits of ranges are inclusive)
- Sex (Male, Female)
- Race (American Indian or Alaska Native, Asian, Black/African American, Native Hawaiian/Other Pacific Islander, White and Other)
- Ethnicity (Hispanic or Latino, Not-Hispanic or Latino)
- Baseline Weight (kg)
- Baseline Height (cm)
- Baseline Body mass index (BMI) (kg/m^2) calculated from weight and height
- BMI category ($\leq$30 kg/m^2, BMI >30 kg/m^2 [defined as obese])
- Systemic corticosteroids use (per Interactive Response Technolgy (IRT))
- SARS-CoV-2 virus result (per CRF) (Positive/Negative) (Note: SARS-CoV-2 infection is laboratory-confirmed by PCR prior to randomization)
- Cycle Threshold, if SARS-CoV-2 positive (per CRF, as available)
- Baseline IL-6 (pg/mL)
- Baseline CRP (mg/L)
- Baseline absolute neutrophil counts
- Baseline Neutrophils-Lymphocyte Ratio (NLR)
- Non-IL6/6R therapy under and EUA at randomization (for Cohorts 2 and 3 only)

Other baseline disease characteristic variables for this study population are as follows.

Pneumonia status at baseline – (selected from Pneumonia Status at Baseline CRF)

- History of or current chronic hypercapnic respiratory failure (yes/no) – (selected from NEWS2 CRF at baseline or Pneumonia status at baseline CRF, depending on original or amended versions of CRFs)
- Presence of pneumonia based on historical chest X-ray or CT scan (yes/no)

- Duration of illness prior to baseline (calculated as earliest onset date for symptoms of pneumonia to first dose date. If date of earliest onset of pneumonia is missing, then date is not imputed)
- Presence of rales/crackles on lung auscultation (yes/no)
- Presence of documented fever in medical record (yes/no)
- Highest temperature recorded 24 hours prior to dosing

  (All temperature is converted to Oral temperature. Rectal/Tympanic minus 0.4 deg C = Oral deg C, or Axillary/Temporal plus 0.4 deg C = Oral deg C)
- Baseline oxygen device type used (in case multiple devices are recorded at baseline, the "worst" device type will be selected)
  - Nasal cannula
  - Face mask
  - Non-rebreather mask
  - High-flow nasal cannula
  - Non-invasive ventilation
  - Invasive ventilation (mechanical ventilation)
  - Extracorporeal life support (eg, ECMO)
- Use of vasopressors (yes/no)
- Use of renal replacement therapy (yes/no)
- Requires treatment in intensive care unit (ICU) (yes/no)

Oxygen administration and Oxygenation at baseline

- SpO_2 % (peripheral capillary oxygen saturation) (range is 0% to 100%)
- FiO_2 (fraction of inspired oxygen) (range is 0.0 to 1.0) – Refer to Phase 2 SAP for derivations
- SpO_2 / FiO_2 ratio– Refer to Phase 2 SAP for derivations

Hospital or ICU stay

- Length of hospital stay including ICU prior to randomization (days)
- Admitted into ICU during hospital stay prior to randomization (yes/no)
- Length of ICU stay prior to randomization (days)

4.2. Medical History

Medical history will be coded to a Preferred Term (PT) and associated primary System Organ Class (SOC) according to Medical Dictionary for Regulatory Activities (MedDRA®) version 23.0.

4.3. Pre-Treatment / Concomitant Medication

Medications/Procedures will be recorded from the day of informed consent until the final study assessment (Day 29 or discharge or death). Medications will be coded to the ATC level 2 (therapeutic main group) and ATC level 4 (chemical/therapeutic subgroup), according to WHO Drug Dictionary (WHODD) version 202003. Patients will be counted once in all ATC categories linked to the medication.

Prior medications/procedures are: medications taken or procedures performed prior to administration of the study drug.

Concomitant medications/procedures are: medications taken or procedures performed following the first dose of study drug through the final study assessment (Day 29 or discharge or death). This includes medications taken that started before the study and are ongoing during the study.

Concomitant medications in a hospitalized population change daily and are difficult to collect in the setting of a pandemic with limited resources in some settings. It is therefore difficult to attribute to success and failure of therapy and impact on safety. Therefore, only select concomitant medications will be captured in this trial. The select list of medications include corticosteroids, remdesivir, lopinavir-ritonavir, chloroquine, hydroxychloroquine, interferon beta, and convalescent serum.

Analysis of medications data will be focused on the targeted medications (specified in the protocol) that are expected to be reviewed and recorded by sites. **Targeted medications** include but are not limited to:

- antipyretics, such as aspirin, acetaminophen, ibuprofen, and other non-steroidal anti inflammatory drugs (NSAIDs)
- warfarin
- cyclosporine A
- theophylline
- digoxin
- antiepileptics, such as carbamazepine (Carbatrol®, Tegretol®), divalproex (Depakote®), phenytoin (Dilantin®), valproic acid (Depakene®);
- antiarrhythmics, such as disopyramide (Norpace®), procainamide (Procan®, Pronestyl®), quinidine (Quinidex®, Quin Release Quin- G®)
- antivirals, antibacterials, and antifungals
- anti-parasitics (chloroquine or hydroxychloroquine)
- interferon beta
- corticosteroids

- convalescent serum
- angiotensin receptor blockers, such as Azilsartan (Edarbi), Candesartan (Atacand), Eprosartan (Teveten), Irbesartan (Avapro), Losartan (Cozaar), Olmesartan (Benicar), Telmisartan (Micardis), Valsartan (Diovan)
- angiotensin converting enzyme inhibitors: benazepril (Lotensin), captopril (Capoten), enalapril (Vasotec), fosinopril (Monopril), lisinopril (Prinivil, Zestril), moexipril (Univasc), perindopril (Aceon), quinapril (Accupril)

4.4. Rescue Medication/or Prohibited Medication During Study

There are no rescue, prohibited or permitted medications in this study, except for those in the exclusion criteria of study enrollment. Patients may continue their normal regimen of medications and procedures. All data collected on medications/procedures (pre-treatment and concomitant) will be summarized.

4.5. Efficacy Variables

In this section, details on the derivation of efficacy variables are provided. Distinction is made between the term "efficacy variable" and "efficacy endpoint".

Whether an efficacy variable is used as a primary efficacy endpoint, key secondary endpoint, or other secondary efficacy endpoint will depend on the population of interest for the variable.

4.5.1. Primary Efficacy Variable

The primary efficacy variable is the **proportion of patients with at least 1-point improvement in clinical status from baseline to Day 22 using the 7-point ordinal scale.**

The **ordinal scale** is an assessment of the clinical status of a patient (Peterson, 2017 [11]). The 7-point ordinal scale is as follows:

1. Death;
2. Hospitalized, requiring invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO);
3. Hospitalized, requiring non-invasive ventilation or high flow oxygen devices;
4. Hospitalized, requiring supplemental oxygen;
5. Hospitalized, not requiring supplemental oxygen – requiring ongoing medical care (COVID-19 related or otherwise)
6. Hospitalized, not requiring supplemental oxygen – no longer requires ongoing medical care
7. Not hospitalized

The clinical status of a patient will be assessed in the morning to consider the worst assessments for the previous day (ie, midnight to midnight; 00:00 – 00:00 [24 hour clock]). If it is the first assessment and 24 hours of data are not available, then the clinical status will be recorded at randomization. Data for clinical status (ordinal scale) is recorded on the *Clinical Status Assessment Ordinal Scale* case report form (CRF).

Baseline clinical status (prior to dosing on Day 1) in the Phase 3 Cohort 1 patient population (randomized and treated) will take values 2, 3, or 4, based on their respiratory status on the baseline pneumonia treatment page, using the value 2 for patients on mechanical ventilation or ECMO, 3 for patients on high-flow nasal canula, or non-mechanical ventilation (BiPaP) and 4 for patients on other less intensive oxygen therapy. Patients who are randomized but not treated, for example who die or are discharged prior to dosing, are excluded from the analysis.

Post-baseline clinical status on a Study Day can take values 1 through 7 on ordinal scale and data will be as recorded on the ordinal scale case report form (CRF). However, since the ordinal scale on a given study is recorded in the morning and it reflects the patient's clinical status during the prior 24 hour period, the value will be attributed to the prior study day. Derivations for death and hospital discharge will be done as described below. Handling of other missing post-baseline values is also mentioned below. After these methods of data handling, any remain missing data will be left missing and not imputed.

Change from baseline on the ordinal scale <in clinical status> on Study Day is defined as post-baseline clinical status on Study Day minus baseline clinical status

Handling of deaths on ordinal scale (post-baseline):

For patients who are recorded on Day X as having died on the 7-point scale, the clinical status ordinal scale value will be retained for the study day X-1 and the clinical status for Study Day X will be derived as "1=Death". After this, the clinical status ordinal value will be carried forward until Day 29 (for key secondary efficacy endpoint [see Section 4.5.2.3]) and until end of study (Day 60) (for other endpoints as specified). Death date may be obtained from *Adverse Events* or *Study Completion* CRFs or phone call at Day 60 (end of study).

For example, a patient's clinical status may be recorded as "2= Hospitalized, requiring invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO)" on the morning of Day 5, but the patient dies later in the evening of Day 5. In this case, clinical status (observed) = 2 for Day 4, and clinical status (derived) = 1 for Day 5 and carried forward for all subsequent study days through Day 20 or Day 60 (as described above).

Handling of discharge on ordinal scale (post-baseline):

Similarly, hospital discharge information may be obtained from *Hospital & ICU Admission & Discharge* CRF. Earliest available date of hospital discharge (Day X) will be used. Clinical status as recorded on the day of discharge (Day X) will be retained as recorded and clinical status for the study day after discharge date (e.g., Day X+1) will be derived as "7=Not hospitalized". After this study day (X+1), the clinical status ordinal value will be carried forward until end of study. If it is learned during follow-up that some patients are subsequently readmitted for COVID-related reasons, the data will be explored to assess effects of readmission in the analyses.

Handling of other missing data on ordinal scale (post-baseline):

For patients who are alive and not discharged (i.e., still hospitalized), missing clinical status (ordinal scale) value on a given study Day X will be imputed using data on the type of oxygen delivery device used by the patient on the previous day, i.e., Day X-1. Clinical status = 2 (if using invasive mechanical ventilation or ECMO), or 3 (if using non-invasive mechanical ventilation, or high-flow nasal cannula), or 4 (if using non-rebreather mask, nasal cannula or

simple face mask) or 5/6 (if not using supplemental oxygen). Clinical status of 5 or 6 will be imputed as the worst value of "5= Hospitalized, not requiring supplemental oxygen – requiring ongoing medical care (COVID-19 related or otherwise)".

Phase 3 Cohort 1:

For each treatment group (sarilumab 400 mg or placebo) in Phase 3 Cohort 1, the primary efficacy endpoint will be calculated as:

$$\frac{\text{Number of patients for whom change from baseline on the ordinal scale is} \geq +1 \text{ at Day 22}}{\text{Number of patients in the treatment group in Phase 3 Cohort 1 critical stratum who were receiving mechanical ventilation at baseline}}$$

Phase 3 Cohort 2:

For each treatment group (sarilumab 800 mg or placebo) in Cohort 2, the primary efficacy endpoint will be calculated as:

$$\frac{\text{Number of patients for whom change from baseline on the ordinal scale is} \geq +1 \text{ at Day 22}}{\text{Number of patients in the treatment group in Phase 3 Cohort 2 (i.e., who received mechanical ventilation at baseline)}}$$

Phase 3 Cohort 3:

For each treatment group (sarilumab 800 mg or placebo) in Cohort 3, the primary efficacy endpoint will be calculated as:

$$\frac{\text{Number of patients for whom change from baseline on the ordinal scale is} \geq +1 \text{ at Day 22}}{\text{Number of patients in the treatment group in Phase 3 Cohort 3 (i.e., who were not on mechanical ventilation but received high intensity oxygen therapy* at baseline)}}$$

*High intensity oxygen therapy will be per IRT stratum (as defined in the protocol).

Note that the numerator and denominator for primary efficacy endpoint could be different depending on the analysis population, i.e., ITT for primary analysis or PPS for sensitivity analysis (See Section 3).

4.5.2. Key Secondary Efficacy Variables

The key secondary efficacy variables for the Phase 3 portion of the study are as follows.

1. Proportion of patients with at least 1-point improvement in clinical status from baseline to Day 22 (Cohort 1 only)
2. Proportion of patients who recover (discharged, or alive without supplemental oxygen use or at pre-COVID oxygen use) by Day 22
3. Proportion of patients who die through Day 29.

However, the key secondary endpoints will be assessed differently depending on the Phase 3 Cohorts. Derivations for variable 1 are in Section 4.5.1 and for variables 2 and 3 are in Section 4.5.2.2 and Section 4.5.2.3.

4.5.2.1. At least 1-point improvement in clinical status – Phase 3 Cohort 1 Critical Stratum only

For each treatment group (sarilumab 400 mg or placebo), the key secondary efficacy endpoint in Phase 3 Cohort 1 only, will be calculated as:

- Number of patients for whom change from baseline on the ordinal scale is ≥+1 at Day 22 divided by number of patients in the treatment group in Phase 3 Cohort 1 Critical stratum (ITT)

This endpoint will be analyzed in the ITT population. Derivation of variables for this endpoint are given in Section 4.5.1.

4.5.2.2. Recovery from COVID-19 – Phase 3 Cohort 1, Cohort 2 and Cohort 3

Baseline Oxygen use is defined as the oxygen delivery device given to the hospitalized patient prior to being dosed and is obtained from the CRF for *Pneumonia Status at Baseline*. Oxygen delivery devices may be (in increasing order of complexity): 1) nasal cannula, 2) face mask, 3) non-rebreather mask, 4) high-flow nasal cannula, 5) non-invasive ventilation (eg, BiPAP), 6) invasive ventilation (mechanical ventilation), or 7) extracorporeal life support (eg, extracorporeal membrane oxygenation [ECMO]). Other types of devices used in this study will be classified into categories 1 through 6 and finalized prior to the first database lock. In case multiple devices are recorded at baseline, then the oxygen device in the highest category will be used. In case of missing data on oxygen devices on the CRF for *Pneumonia Status at Baseline*, the CRF for *Oxygen Administration and Oxygenation* at Baseline will be used to impute this data.

Post-baseline Oxygen use on a Study Day is defined as the oxygen delivery device used on Study Day as recorded on the *Oxygen Administration and Oxygenation* CRF on that Study Day. If oxygen delivery device is blank or site records that no supplemental oxygen was used, then post-baseline oxygen use will be derived as "None".

A patient is considered recovered if they are either discharged from the hospital, or are alive in the hospital without supplemental oxygen use or at pre-COVID oxygen use.

Patients who recover at Study Day are defined as those who are:

1. **discharged**, or

   (Subject discharged from hospital before Day 29 is marked as Yes, or Hospital discharge date is non-missing on the *Hospital & ICU Admission & Discharge* CRF)

2. **alive without supplemental oxygen use**, or

   (The question "Was any supplemental Oxygen or mechanical ventilation used?" is marked as "No" on the *Oxygen Administration and Oxygenation* CRF), or

3. **at pre-COVID oxygen use**.

(Post-baseline oxygen use at Study Day is the same as or better than the baseline oxygen use category.)

Phase 3 Cohort 1:

For each treatment group (sarilumab 400 mg or placebo) in Phase 3 Cohort 1, the key secondary efficacy endpoint will be calculated as:

- Number of patients who recovered divided by number of patients in the treatment group in Phase 3 Cohort 1 Critical stratum receiving mechanical ventilation without ECMO at baseline

Phase 3 Cohort 2 and Cohort 3:

For each treatment group (sarilumab 800 mg or placebo) in Phase 3 Cohort 2 and Cohort 3, the key secondary efficacy endpoints will be calculated as:

- Number of patients who recovered divided by number of patients in the treatment group in Phase 3 Cohort 2 receiving mechanical ventilation at baseline
- Number of patients who recovered divided by number of patients in the treatment group in Phase 3 Cohort 3 on high-intensity oxygen therapy without mechanical ventilation at baseline

4.5.2.3. Mortality (Survival Status)

A patient's survival status at Study Day X is derived as "Died" if there is death data on that study day on any of these CRFs – *Study Completion, Adverse Events, Phone Call, or Subject Eligibility* (in case of screen failures or randomized not treated). In case of missing data on these CRFs, the *Clinical Status Assessment Ordinal Scale* CRF data may be used to check death status. However, this CRF could have a lag of upto 24 hours and if clinical status ordinal scale is recorded as 1= "Death" on Day X, then death date will be assigned to the previous Day X-1.

Otherwise, the patient's survival status is "Alive" on that Study Day. If a patient dies, survival status of death will be carried forward after the study day that patient dies.

Phase 3 Cohort 1:

For each treatment group (sarilumab 400 mg or placebo) in Phase 3 Cohort 1, the key secondary efficacy endpoint will be calculated as:

- Number of patients who died through Day 29 divided by number of patients in the treatment group in Phase 3 Cohort 1 Critical stratum receiving mechanical ventilation without ECMO at baseline.
- Number of patients who died through Day 29 divided by number of patients in the treatment group in Phase 3 Cohort 1 ALL Critical stratum (ITT).

Phase 3 Cohort 2 and Cohort 3:

For each treatment group (sarilumab 800 mg or placebo) in Phase 3 Cohort 2 and Cohort 3, the key secondary efficacy endpoints will be calculated as:

- Number of patients who died through Day 29 divided by number of patients in the treatment group in Phase 3 Cohort 2 receiving mechanical ventilation at baseline
- Number of patients who died through Day 29 divided by number of patients in the treatment group in Phase 3 Cohort 3 on high-intensity oxygen therapy without mechanical ventilation at baseline

4.5.3. Other Secondary Efficacy Variables

Additional secondary efficacy variables and populations of interest specified in the protocol are shown in Table 2.

Table 2: Phase 3 – Other Secondary Efficacy Endpoints

Efficacy Variable	**Populations of Interest**				
	Cohort 1: sarilumab 400 mg or placebo			**Cohort 2: sarilumab 800 mg or placebo** (receiving mechanical ventilation at baseline)	**Cohort 3: sarilumab 800 mg or placebo** (receiving high-intensity oxygen therapy without mechanical ventilation at baseline)
	Critical ITT	**Critical receiving mechanical ventilation at baseline**	**Critical not receiving mechanical ventilation at baseline[1]**		
Proportion of patients alive not requiring invasive mechanical ventilation or ECMO at Day 22	Secondary	(Same as Primary)	Descriptive	(Same as Primary)	Secondary
Proportion of patients receiving invasive mechanical ventilation or ECMO at Day 22	Secondary	Secondary	Descriptive	Secondary	Secondary
Proportion of patients with a 2-point improvement in clinical status on the 7-point ordinal scale from baseline to Day 22	Secondary	Secondary	Descriptive	Secondary	Secondary
Time to at least 1-point improvement in clinical status assessment from baseline on the 7-point ordinal scale	Secondary	Secondary	Descriptive	Secondary	Secondary

Efficacy Variable	Populations of Interest				
	Cohort 1: sarilumab 400 mg or placebo			Cohort 2: sarilumab 800 mg or placebo (receiving mechanical ventilation at baseline)	Cohort 3: sarilumab 800 mg or placebo (receiving high-intensity oxygen therapy without mechanical ventilation at baseline)
	Critical ITT	Critical receiving mechanical ventilation at baseline	Critical not receiving mechanical ventilation at baseline[1]		
Time to at least 2-point improvement in clinical status assessment from baseline on the 7-point ordinal scale	Secondary	Secondary	Descriptive	Secondary	Secondary
Proportion of patients discharged and alive at Day 22	Secondary	Secondary	Descriptive	Secondary	Secondary
Time to recovery (discharged, or alive without supplemental oxygen use or at pre-COVID oxygen use)	Secondary	Secondary	Descriptive	Secondary	Secondary
Time to death (all-cause mortality)	Secondary	Secondary	Descriptive	Secondary	Secondary
Number of ventilator free days between baseline and Day 8, 15, 22, and 29	Secondary	Secondary	Descriptive	Secondary	Secondary
Number of days of hospitalization among survivors up to Day 8, 15, 22, and 29	Secondary	Secondary	Descriptive	Secondary	Secondary

[1] The main approach in Cohort 1 to the analysis of patients "Critical not receiving mechanical ventilation at baseline" will be through an analysis of a subset of the overall ITT population to assess if an effect observed in the ITT population is observed both in the subset of patients on mechanical ventilation and not on mechanical ventilation at baseline.

4.5.3.1. Mechanical ventilation or ECMO related outcomes

Patients are considered to be receiving invasive mechanical ventilation or ECMO, if their post-baseline oxygen use (see Section 4.5.2.2) is invasive (mechanical) ventilation or extracorporeal life support. Otherwise, patients are considered as not requiring invasive mechanical ventilation or ECMO.

Proportion of patients alive not requiring mechanical ventilation or ECMO at Day 22, is calculated as number of patients who are (alive and not on mechanical ventilation or ECMO) at Day 22 divided by the population of interest.

Proportion of patients receiving mechanical ventilation or ECMO at Day 22 is calculated as number of patients who are receiving mechanical ventilation or ECMO at Day 22 divided by the population of interest.

4.5.3.2. At least x-point improvement in clinical status on ordinal scale related outcomes

Proportion of patients with a 2-point improvement in clinical status on the 7-point ordinal scale from baseline to Day 22 is calculated as number of patients for whom change from baseline on the ordinal scale is ≥+2 at Day 22 divided by population of interest.

Time to at least 1 (2)-point improvement (event) in clinical status on the 7-point ordinal scale is defined as the first date of change from baseline in clinical status ≥+1(+2) minus first dose date. Patients who do not experience improvement of 1 (2) point(s) or more on the ordinal scale will be censored at the last observed time point. Patients who die at any time in the study will be censored at Day 60 (end of study – phone call follow-up) in case of final database lock (In case of an interim data lock, censoring will occur on the study day of the longest follow-up of an alive patient at the time of data cut-off, e.g., if the longest follow-up of an alive patient at the data cut-off date of interim lock is Day 15, then patients who died will be censored at Day 15.)

4.5.3.3. Discharged and Alive

Proportion of patients who are discharged and alive at Day 22 is the number of patients who are alive and have a discharge date (or marked as discharged from hospital before Day 29 in case date is missing) on the *Hospital & ICU Admission & Discharge* CRF divided by the population of interest.

4.5.3.4. Time-to-recovery

Time to recovery (event) is computed as the first date a patient has recovered (as defined in Section 4.5.2.2) minus the first dose date + 1. Patients who do not recover will be censored at the last observed time point. Patients who die at any time in the study will be censored at Day 60 (end of study – phone call follow-up) in case of final database lock (In case of an interim data lock, censoring will occur on the study day of the longest follow-up of an alive patient at the time of data cut-off, e.g., if the longest follow-up of an alive patient at the data cut-off date of interim lock is Day 15, then patients who died will be censored at Day 15.)

4.5.3.5. Time-to-death (all-cause mortality)

Time to death (all-cause mortality) (event) will be computed as the date of death (death for any cause) minus the first dose date +1. Patients who are alive will be censored at the last observed timepoint. Survival status is derived from the CRFs as described in Section 4.5.2.3. If the date of death is missing (even after data query to the study site) but patient is marked as died, then death date will be imputed as Study Day 1.

4.5.3.6. Ventilator-free days

Assisted ventilation is defined as the use of invasive mechanical ventilation or extracorporeal life support (eg, extracorporeal membrane oxygenation [ECMO]). Duration of ventilation and ventilator-free days variables are defined in this section.

Ventilator-free days based on Study Day 29

For those receiving assisted ventilation at baseline see derivations below:

- The **duration of ventilation** is defined as [Last date of assisted ventilation in the hospital – *max*(first date of assisted ventilation, first dose date)], if last day is prior to Day 29. (Note: Some patients may be randomized while on assisted ventilation. Hence maximum of first date of assisted ventilation and first dose date is used.) Otherwise, **duration of ventilation** is (Date of Study Day 29 - first date of assisted ventilation).

For those not receiving assisted ventilation at baseline see derivations below:

- In patients who never require assisted ventilation, **duration of ventilation** is zero.
- In patients who initiated assisted ventilation during the study, **duration of ventilation** is defined as (stop date of assisted ventilation minus start date of assisted ventilation). If a patient has been on and off ventilation then it will be the summation of all days on ventilation. If the assisted ventilation is ongoing and there is no stop date, then use the last observed date on assisted ventilation.

Ventilator-free days (VFD) depend on both duration of ventilation and mortality through study Day 29. The maximum number of VFD can be 28 days. (Assuming that patient is observed at the same time on each study day, there are 28 24-hour periods from Day 1 to Day 29.)

VFD will be calculated and reported in patients using assisted ventilation (or not) at baseline. VFD may also be analyzed in these 2 groups after excluding patients using ECMO at baseline.

Ventilator-free days are defined as 28 minus duration of ventilation, with following considerations.

- Patients who do not survive 29 days will be assigned zero VFD.
- Patients discharged and alive prior to Day 29 on unassisted breathing will be assumed to remain on unassisted breathing through Day 29.
- For patients who experience multiple episodes of assisted ventilation (e.g. may be on a ventilator, then come off the ventilator and be on a ventilator again) ventilator free days will be computed between each episode of ventilation and total VFD calculated.

Ventilator-free days based on Study Days 8, 15, 22

VFD variable for each patient will also be calculated similarly (as above) based on data up to Study Day 8, 15, and 22 (replace Day 29 in the above derivation rules by respective study day). Maximum VFD will be 7 days, 14 days, and 21 days, respectively.

4.5.3.7. Days of Hospitalization

Length of hospitalization after randomization will be calculated as date of the last observed study day in the hospital minus first dose date +1. If a patient dies, then date of last observed study day will be death date. If a patient is discharged, then date of last observed study day will be the date of discharge.

4.6. Safety Variables

4.6.1. Adverse Events and Serious Adverse Events

Serious adverse events and AESIs will be collected from the time of informed consent signature and then at each visit through Day 29 or discharge or death, whichever is sooner. Patients discharged prior to Day 29 will have a follow-up phone call on Day 29 to assess vital status, collect data on serious adverse event and history of hospital re-admission. At the Day 60 phone call follow-up, data on vital status of the patient (alive or dead) and history of hospital re-admission will be collected. All adverse events are to be coded to a "Preferred Term (PT)" and associated primary "System Organ Class (SOC)" according to the Medical Dictionary for Regulatory Activities (MedDRA version 23.0).

An Adverse Event is any untoward medical occurrence in a patient administered a study drug which may or may not have a causal relationship with the study drug.

A Serious Adverse Event is an adverse event (AE) that is classified as serious according to the criteria specified in the protocol.

Infusion reactions are defined as any relevant AE that occurs during the infusion or within 2 hours after the infusion is completed.

The severity of AEs will be graded using the NCI-CTCAE v5. Adverse events not listed in the NCI-CTCAE v5 will be graded according to the following scale:

Table 3: Grading System for Adverse Events Not Listed in NCI-CTCAE

Grade	Severity	Description
1	Mild	Asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
2	Moderate	Minimal, local, or noninvasive intervention indicated; limiting age-appropriate instrumental ADL*.
3	Severe	Severe or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care ADL**.
4	Life–threatening	Life-threatening consequences; urgent intervention indicated.
5	Death	Death related to AE

* Instrumental Activities of Daily Living (ADL) refer to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.

** Self-care ADL refer to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden.

Laboratory results, vital signs, or ECG (if feasible) abnormalities will be recorded as AEs if they are medically relevant: symptomatic, requiring corrective therapy, leading to treatment discontinuation and/or fulfilling a seriousness criterion.

The **secondary safety variables** are (also see Section 4.6.2):

1. Proportion of patients with serious adverse events
2. Proportion of patients with Grade 4 neutropenia (absolute neutrophil count (ANC) < 500/mm^3)
3. Proportion of patients with severe or life-threatening bacterial, invasive fungal, or opportunistic infection
4. Proportion of patients with severe or life-threatening bacterial, invasive fungal, or opportunistic infection in patients with Grade 4 neutropenia (ANC < 500/mm^3)
5. Proportion of patients with hypersensitivity reactions
6. Proportion of patients with infusion reactions
7. Proportion of patients with gastrointestinal perforation
8. White cell count, hemoglobin, platelets, creatinine, total bilirubin, ALT, AST, on Days 1, 3, 5, 8, 11, 15, and 29 (if still hospitalized)

4.6.2. Adverse Events of Special Interest

Adverse events of special interest (AESI) are AEs (serious or non-serious) of scientific and medical interest specific to this drug program, for which ongoing monitoring and rapid communication by the investigator to the sponsor will be done.

In this study, the AESIs are listed below:

- Grade 4 neutropenia (absolute neutrophil count (ANC) < 500/mm^3) :
  - selected from Local Lab-Hematology CRF
- Grade 4 neutropenia (ANC < 500/mm^3) with concurrent severe or life-threatening bacterial, invasive fungal, or opportunistic infection :
  - selected from Adverse Event CRF
- Grade ≥2 infusion related reactions
  - selected from Adverse Events CRF
- Grade ≥2 hypersensitivity reactions
  - selected from Adverse Events CRF
- Increase in alanine transaminase (ALT) or aspartate aminotransferase (AST) ≥ 3 X upper limit of normal (ULN) (for patients with normal baseline) or > 3 X ULN AND at least 2 fold increase from baseline value (for patients with abnormal baseline)
  - selected from Local Lab-Chemistry CRFs

- Invasive bacterial or fungal infections of clinical significance with confirmed diagnosis based on the investigator's assessment with appropriate diagnostic workups and consultations
  - selected from Adverse Event CRF

4.6.3. Laboratory Safety Variables

The clinical laboratory data consists of blood chemistry (including C-Reactive Protein, liver function tests, creatinine and other), hematology, urinalysis, infection testing, SARS-Cov-2 RT-PCR, and other (as specified in the protocol).

Clinical laboratory values will be converted to standard international (SI) units and grouped by function in summary tables. Conventional unit may be provided. Functions are defined as follows:

- Liver function including ALT, AST, alkaline phosphatase, total bilirubin,
- Renal function including creatinine, uric acid,
- Electrolytes including sodium, potassium,
- C-Reactive Protein (CRP),
- Creatine Phosphokinase (CPK)
- Metabolic parameters including total proteins, albumin,
- White blood cells (WBCs) including WBCs count and differential count (neutrophils, lymphocytes, eosinophils, basophils, monocytes),
- Red blood cells (RBCs) and platelets including red blood cells count, hemoglobin, hematocrit and platelets count,
- Other

4.6.4. Vital Signs

Vital signs, including temperature, blood pressure, pulse, and respiration, are recorded at multiple time points according to Schedule of Time and Events table (See Section 10.1).

4.6.5. Physical Examination Variables

A targeted physical examination including lung auscultation will be performed at time point according to Schedule of Time and Events table (See Section 10.1). Care will betaken by the investigator to examine and assess any abnormalities that may be present, as indicated by the patient’s medical history.

During screening period, if any existing clinically significant abnormalities are present, these will be recorded in the *Medical History* CRF. Post-screening period, if any new clinically significant abnormalities are present (per investigator discretion), the relevant event will be recorded in the *Adverse Event* CRF, if applicable.

4.7. Pharmacokinetic Variables

The PK variables are the concentration of sarilumab and concentration of sIL-6R in serum at each time point specified in the Schedule of Time and Events table (See Section 10.1).

4.8. Pharmacodynamic and Other Biomarker Variables

Exploratory endpoint variables include measurement of SARS-CoV-2 in OP or NP swabs over time using RT-PCR. Qualitative (positive or negative) or relative quantitation of viral copies may be evaluated. Pharmacodynamic variables may include the time to reach a negative OP or NP RT PCR result.

Additional biomarker testing may include, but not be limited to, evaluation of inflammatory cytokines in serum, and ANC.

Pharmacodynamic variables include the ANC, and the concentration of IL-6 in serum at each time point.

These results may be reported outside the clinical study report (CSR).

5. STATISTICAL METHODS

For continuous variables, descriptive statistics will include the following: the number of patients reflected in the calculation (n), mean, median, standard deviation, quartiles, minimum, and maximum.

For categorical or ordinal data, frequencies and percentages will be displayed for each category.

Separate analyses will be conducted for Phase 3 Cohort 1, Cohort 2, and Cohort 3. The main focus of the Phase 3 analysis will be on patients in

1. Cohort 1† Critical stratum randomized to sarilumab 400 mg IV or placebo treatment groups,
2. Cohort 2 randomized to sarilumab 800 mg IV or placebo treatment groups receiving mechanical ventilation at baseline, and
3. Cohort 3 randomized to sarilumab 800 mg IV or placebo treatment groups receiving high-intensity oxygen therapy without mechanical ventilation at baseline.

† Note that Cohort 1 consists of all Phase 3 patients randomized to sarilumab 400 mg IV, sarilumab 200 mg IV or placebo across disease severity strata. Within this Cohort, the primary focus of this SAP is on the efficacy and safety data on patients in the critical stratum randomized to sarilumab 400 mg or placebo. For patients in the critical stratum randomized to sarilumab 200 mg, only safety data will be analyzed.

Efficacy analyses of previously enrolled Phase 3 Cohort 1 patients randomized to sarilumab 200 mg versus placebo in all disease severity strata will be done for exploratory purpose only. Efficacy and safety analyses of previously enrolled Phase 3 Cohort 1 patients randomized to sarilumab 400 mg IV versus placebo for the disease severity strata of severe, MSOD, or immunocompromised, will also be done for exploratory purpose only. These exploratory analyses will be designed similarly as the main analyses described in this SAP.

5.1. Demographics and Baseline Characteristics

Demographic and baseline characteristics variables given in Section 4.1 will be summarized descriptively by treatment group, and all groups combined. These will be analyzed for the ITT population and the subset population of interest within ITT as specified by the primary efficacy endpoints.

5.2. Medical History

Medical history will be summarized by SOC and PT and by treatment group and all groups combined in the ITT population and the subset population of interest within ITT as specified by the primary efficacy endpoints.

5.3. Prior/concomitant Illnesses and Medications

Prior or concomitant medications/procedures will be summarized by treatment groups. Focus of the results will be on the list of targeted medications (Section 4.3) in the ITT population.

5.4. Rescue/ Prohibited Medications if applicable

Not applicable. See Section 4.4.

5.5. Subject Disposition

The following summaries will be provided for all screened patients.

Total number of

- Screened patients: met the inclusion criteria regarding the target indication and signed the ICF
- Randomized patients: received a randomization number per IRT
- Patients treated but not randomized
- Patients randomized and not treated
- Screen failures
- Reasons for screen failure

The following summaries will also be provided for the ITT population (see Section 3.1 for definition).

- The total number of randomized patients
- The total number of patients who completed the study
- The total number of patients who discontinued the study, and the reasons for discontinuation
- The total number of patients who discontinued from study treatment, and the reasons for discontinuation.

  (Note: Until Protocol Amendment #3 was effective, patients were given only a single dose. After Protocol Amendment #4, repeat dosing is permitted after 24 hours of dosing and weekly under protocol-defined criteria. Discontinuation from study treatment refers to either study treatment interruption (if single dose) or study drug discontinuation (if multiple doses).)
- A listing of patients treated but not randomized, patients randomized but not treated, and patients randomized but not treated as randomized
- A listing of patients prematurely discontinued from treatment, along with reasons for discontinuation

  (Note: Applicable only on data after Protocol Amendment #4 is in effect, as described above.)

In addition, the number of patients in each treatment group and combined groups will be summarized for the analysis sets of ITT population and Per Protocol Set defined in Section 3.1 and Section 3.2, Safety population defined in Section 3.3, and Pharmacokinetics Analysis Set defined in Section 3.6. This analysis will be done in ITT.

5.6. Extent of Study Treatment Exposure and Compliance

Prior to Protocol Amendment #4 treated patients were given a single dose of study drug (sarilumab 400 mg IV, sarilumab 200 mg IV, or placebo IV). After Protocol Amendment #4, repeat dosing of study drug was permitted after 24 hours (Day 2) if no clinical response is observed and additional (repeat) weekly dosing is permitted (under protocol-defined criteria). In Cohort 1, a maximum of 6 doses of study drug in total are permitted during the study. In Cohort 2, a maximum of 4 doses of study are permitted during the study with no dose after Day 21.

5.6.1. Measurement of Compliance

Treatment compliance in a given patient is defined as the number of fully completed infusions of study drug divided by number of doses administered (applicable, both, to patients receiving only single dose or multiple doses since Protocol Amendment #4). Treatment compliance will be summarized by treatment group using descriptive statistics based on the Safety population.

5.6.2. Exposure to Investigational Product

Exposure to study drug will be examined for each patient as recorded on the *Study Drug Administration-IV* and *Study Drug Administration-IV (Re-Dose)* CRFs.

The following variables will be descriptively summarized (and/or) provided in patient listings by treatment group:

- Number of patients who received single dose
- Number of patients who received multiple doses
- Number of patients receiving 1, 2, 3, 4, 5, or 6 doses (up to maximum if maximum is higher than 6, even though 6 is maximum specified in protocol)
- Duration of intravenous infusion (mins) (infusion end time – infusion start time) on the days of dosing
- Location of administration on the days of dosing
- Total volume of drug administered (units: mL) on the days of dosing
- Number of patients with infusion interruptions on days of dosing
- Number of patients receving study drug over time, e.g., on Day 1, Day 2, and weekly thereafter (Days 8, 15, 22, 29). Cohort 1 can receive maximum of 6 doses. Cohort 2 can receive maximum of 4 doses with no dosing after Day 21.

In addition, duration on study (days) will also be descriptively summarized by treatment group.

5.7. Analyses of Efficacy Variables

Post adaptations to the Phase 3 study design, the main focus of the efficacy analyses will be 3 separate pairwise comparisons, namely in:

1. Cohort 1 critical patients with mechanical ventilation without ECMO at baseline comparing sarilumab 400 mg IV versus placebo,
2. Cohort 2 patients with mechanical ventilation at baseline comparing sarilumab 800 mg IV versus placebo, and
3. Cohort 3 patients without mechanical ventilation but on high-intensity oxygen therapy at baseline comparing sarilumab 800 mg IV versus placebo.

5.7.1. Analysis of Primary Efficacy Variables

The primary efficacy analysis for each comparison will be using the efficacy variable of proportion of patients with at least 1 point improvement in clinical status from baseline to Day 22 using the 7-point ordinal scale (defined in Section 4.5.1). The primary analysis population will be ITT (defined in Section 3.1).

Hypothesis tests of superiority of sarilumab (400 mg or 800 mg dose) versus placebo will be done using the stratified Cochran-Mantel-Haenszel (CMH) test for two proportions. For Cohort 1, the stratification factor will be use of steroids at baseline (yes/no). For Cohort 2 and Cohort 3, the stratification factors will be use of non-IL6/6R therapy under an EUA at randomization (yes/no) and use of steroids at baseline (yes/no).

Estimation of the treatment effect will be provided as differences in proportions and confidence intervals calculated using thestrata-adjusted confidence intervals from CMH method (Zhang, 2016 [13]). Stratification factors are as mentioned above.

For each cohort, p-values and confidence intervals will be reported with overall Type 1 error controlled at 0.05 (2-sided) (see Section 7).

Sensitivity analysis:

Sensitivity analysis will be conducted on the primary efficacy variable using the population of PPS (defined in Section 3.2)

Supportive analyses:

The analysis of the secondary efficacy variable of time to at least 1 point improvement in clinical status from baseline on the 7-point ordinal scale will be used as supportive analysis. Kaplan-Meier estimates on proportion of patients with at least 1 point improvement at Day 22 in each treatment group will be reported for descriptive purpose. Comparisons between groups will be made using Restricted Mean Survival Time (RMST) computed at Day 22 (Kim, 2017 [9]). Two-sided confidence intervals and p-values for specified significance level can be computed for differences in RMST based on asymptotic normal distribution.

Subgroup analyses:

Descriptive analyses will be performed on the primary endpoint to summarize the treatment effects by the following variables:

- Age (<median vs. ≥median) (years)
- Sex (Male vs. Female)
- Race (e.g., White vs. Asian vs. Other)
- Baseline comorbidities (Hypertension, Diabetes, Obesity)
- Steroid use (yes/no)
- Duration of illness (<median vs. ≥median) (days)
- Length of hospital stay include ICU prior to randomization (1-2, 3-4, >=5 days) (days)
- Length of hospital stay including ICU prior to randomization (<median vs. ≥median) (days)
- Length of ICU stay prior to randomization (1-2, 3-4, >=5 days) (days)
- Length of ICU stay prior to randomization (<median vs. ≥median) (days)
- Baseline IL-6 level (<230 pg/mL vs. ≥230 pg/mL; other cutpoints eg, 70 and 140 pg/mL will also be analyzed)
- Baseline CRP tertiles
- Baseline neutrophil-to-lymphocyte ratio tertiles
- Use of other COVID-19 therapies

5.7.2. Analysis of Secondary Efficacy Variables

The statistical methods used for the primary and sensitivity analysis (ITT and PPS) of key secondary efficacy variables will be same as the methods described for the primary efficacy variable.

These key secondary endpoints for the Phase 3 portion of the study will be tested sequentially in a hierarchical manner, while preserving the overall significance level at 0.05 (2-sided).

Cohort 1

1. Proportion of patients with at least 1-point improvement in clinical status assessment from baseline to Day 22 in patients with critical COVID-19
2. Proportion of patients who recover (discharged, or alive without supplemental oxygen use or at pre-COVID oxygen use) by Day 22 in patients with critical COVID-19 receiving mechanical ventilation without ECMO at baseline
3. Proportion of patients who die through Day 29 in patients with critical COVID-19 receiving mechanical ventilation without ECMO at baseline
4. Proportion of patients who die through Day 29 in patients with critical COVID-19

Cohort 2

1. Proportion of patients who recover (discharged, or alive without supplemental oxygen use or at pre-COVID oxygen use) by Day 22 in patients with COVID-19 receiving mechanical ventilation at baseline
2. Proportion of patients who die through Day 29 in patients with COVID-19 receiving mechanical ventilation at baseline

Cohort 3

1. Proportion of patients who recover (discharged, or alive without supplemental oxygen use or at pre-COVID oxygen use) by Day 22 in patients receiving high-intensity oxygen therapy without mechanical ventilation at baseline
2. Proportion of patients who die through Day 29 in patients high-intensity oxygen therapy without mechanical ventilation at baseline

5.7.3. Analysis of Other Secondary Efficacy

Other secondary efficacy variables listed in Section 4.5.3 will be analyzed as follows depending on the variable:

1. Categorical data will be summarized as proportions by Study Day in each treatment group. For the specific time points (e.g., Day 22) in the secondary endpoint, p-values and 95% confidence intervals will be reported for descriptive purpose using stratified CMH test and strata-adjusted CMH method, respectively (as mentioned for the primary endpoints).
2. Differences in time-to-event endpoints by treatment (eg, time to a one category improvement in ordinal scale, and other time-to-event endpoints) will be summarized with Kaplan-Meier estimates and nominal 95% confidence intervals on medians will be reported. P-values and confidence intervals for hazard ratios comparing treatment groups will also be reported using Cox proportional hazards model with stratification factors mentioned earlier. Cumulative incidence rates will be plotted with comparisons between groups descriptively tested using log-rank test.
3. Duration of event (eg, ventilator-free days, days of hospitalization) will be summarized through descriptive statistics including mean, median, and median days with quartiles. P-values using two-sample t-test and 95% confidence intervals using normal approximation will also be reported for descriptive purpose.
4. Change in ordinal scale at specific time points will be summarized by proportions (eg, proportion who have a 1-, 2-, 3-, 4-, or 5-point improvement or 1-, 2-, or 3-point worsening).

5.7.4. Adjustment for Multiple Comparison

Primary analyses in Cohort 1, Cohort 2 and Cohort 3 will be conducted separately. Overall type 1 error will be controlled within each cohort at 0.05 (2-sided) level.

Multiplicity adjustment will be made for interim looks (Section 7) and significance level will be adjusted for final analysis of primary endpoint to control overall type I error. Based on the alpha spending rules for interim analysis, the final significance level for primary endpoint will be 0.0493 (2-sided) .

Multiplicity will be controlled for testing the specified primary ane key secondary endpoints using a hierarchical testing order as shown in Table 4.

Table 4: Testing Strategy for Control of Multiplicity in Phase 3 Cohorts 1, 2 and 3 (Overall Type 1 Error in each cohort will be 0.05 [2-sided])

Order	Timing of Analysis	Type of Endpoint	Endpoint	Population	Example Nominal Significance level, α
Cohort 1					
1	Final	Primary	Proportion of patients with a 1- point improvement in clinical status from baseline to Day 22	Critical COVID-19 patients receiving mechanical ventilation at baseline	0.049†
2	Final	Key Secondary	Proportion of patients with a 1- point improvement in clinical status assessment from baseline to Day 22	All Critical stratum COVID-19 patients (including ventilated or ECMO or not ventilated at baseline)	0.049†
3	Final	Key Secondary	Proportion of patients who recover (discharged, or alive without supplemental oxygen use) by Day 22	Critical COVID-19 patients receiving mechanical ventilation at baseline	0.049†
4	Final	Key Secondary	All-cause mortality at Day 29	Critical COVID-19 patients receiving mechanical ventilation at baseline	0.049†
5	Final	Key Secondary	All-cause mortality at Day 29	All Critical stratum COVID-19 patients (including ventilated or ECMO or not ventilated at baseline)	0.049†
Cohort 2					
1	Final	Primary	Proportion of patients with a 1- point improvement in clinical status from baseline to Day 22	COVID-19 patients receiving mechanical ventilation at baseline	0.049‡
2	Final	Key Secondary	Proportion of patients who recover (discharged, or alive without supplemental oxygen use) by Day 22	COVID-19 patients receiving mechanical ventilation at baseline	0.049‡

Order	Timing of Analysis	Type of Endpoint	Endpoint	Population	Example Nominal Significance level, α
3	Final	Key Secondary	All-cause mortality at Day 29	COVID-19 patients receiving mechanical ventilation at baseline	0.049‡
Cohort 3					
1	Final	Primary	Proportion of patients with a 1- point improvement in clinical status from baseline to Day 22	COVID-19 patients not receiving mechanical ventilation but using high intensity oxygen therapy at baseline	0.049‡
2	Final	Key Secondary	Proportion of patients who recover (discharged, or alive without supplemental oxygen use) by Day 22	COVID-19 patients not receiving mechanical ventilation but using high intensity oxygen therapy at baseline	0.049‡
3	Final	Key Secondary	All-cause mortality at Day 29	COVID-19 patients not receiving mechanical ventilation but using high intensity oxygen therapy at baseline	0.049‡

† If the interim analysis results for primary endpoint are statistically significant, then the Sponsor may decide to stop the cohort earlier and the alpha used at interim will be reallocated to test the key secondary endpoints at the full α=0.05 (2-sided). If the interim analysis on the primary endpoint is not significant, then the Sponsor will continue the follow-up of the cohort and any results on the secondary endpoints at interim will be exploratory. The Sponsor may present nominal p-values for key secondary endpoints at the interim without formal testing at interim.

‡ Interim analyses are planned for Cohort 2 and 3. In case of interim analyses, similar alpha spending and alpha re-allocation rules will be used for Cohorts 2 and 3.

5.8. Analysis of Safety Data

The analysis of safety data will be performed on the SAF, as defined in Section 3.3.

The safety analysis will be based on the reported SAEs and AESIs and other safety information (clinical laboratory evaluations and vital signs).

Thresholds for Potential Clinically Significant Values (PCSV) in laboratory variables, vital signs and ECG are defined in Appendix 10.2 .

The summary of safety results will be presented for each treatment group.

5.8.1. Adverse Events

The verbatim text, the PT, and the primary SOC will be listed in subject listings. Summaries that include frequencies and proportions of patients reporting AEs will include the PTs and the SOCs.

Period of observation: The observation period will be the on-treatment period defined as the day from first dose of study drug (Day 1) to study Day 29. (Patients who are discharged prior to

Day 29 will receive a follow-up phone call to collect data on SAEs (if any), survival and history of hospital re-admission.)

Treatment-emergent AEs (TEAEs) are defined as those AEs that are not present at baseline or represent the exacerbation of a pre-existing condition during the on-treatment period. In this study, only SAEs and AESIs are collected. As such TEAEs will reflect this data.

For details on handling missing data and partial dates, see Section 6.

Summaries of AE incidence in each treatment group will include:

- Overview of TEAEs, summarizing number of events, summarizing number and percentage of patients within the specified category
  - Total number of TEAEs, SAEs, AESIs, serious AESIs
  - Patients with any TEAEs, any SAEs, any AESIs, serious AESIs
  - Patients with any TEAEs leading to study drug interruption or study discontinuation, leading to death
- TEAEs by system organ class (SOC) and preferred terms (PT)
  - All TEAEs
  - TEAEs by relationship to treatment (related, not related)
  - TEAEs by CTC grade (according to the grading scale outlined in Section 4.6.1), presented by SOC and PT
  - TEAEs leading to study drug interruption or study discontinuation
  - TEAEs leading to death
- AESIs by PT
- SAEs by SOC and PT
- Deaths

Counts will be provided according to treatment group for each PT within each SOC. Percentages will be calculated using the number of patients from the safety population in each treatment group.

Primary SOCs will be sorted according to decreasing order of frequency in the combined treatment groups. Within each primary SOC, PTs will be sorted by decreasing frequency of investigational product.

A second type of table with counts of each PT in decreasing order of frequency will also be provided.

5.8.2. Analysis of Adverse Events of Special Interest

Summaries of AESIs (given in Section 4.6.2) will include the following and presented by PT in each treatment group:

1. Incidence of Grade 4 neutropenia (ANC < 500/mm^3)

2. Incidence of severe or life-threatening bacterial, invasive fungal, or opportunistic infection (presented in listings and summarized through patient narratives)
3. Incidence of Grade 4 neutropenia (ANC < 500/mm^3) with concurrent severe or life-threatening bacterial, invasive fungal, or opportunistic infection
4. Incidence of Grade ≥2 hypersensitivity reactions, Grade ≥ 2 infusion reactions, gastrointestinal perforation
5. White cell count, hemoglobin, platelets, creatinine, total bilirubin, ALT, AST, on all study days (available data)
6. Increase in alanine transaminase (ALT) or aspartate aminotransferase (AST) ≥ 3xupper limit of normal (ULN) (for patients with normal baseline) or >3X ULN AND at least 2 fold increase from baseline value (for patients with abnormal baseline)
7. Incidence of any invasive bacterial or fungal infections

5.8.3. Clinical Laboratory Measurements

A treatment-emergent Potential Clinically Significant abnormal value (PCSV) is a laboratory value that was normal at Screening and Baseline but abnormal after treatment with study drug, or a laboratory value that was abnormal at Baseline and exacerbates after treatment with study drug. "Exacerbations" will be identified by the Medical Monitor using clinical judgment. Treatment-Emergent Potentially clinically significant values (PCSVs) will be summarized by treatment group. Additionally, shift table for PCSV may be displayed. See Appendix Section 10.2 for the criteria of PCSV values.

Baseline clinical laboratory analytes and change from Baseline in clinical laboratory analytes to each scheduled assessment time will be summarized with descriptive statistics. Summary statistics will include the number of patients, mean, median, standard deviation, quartiles, minimum, and maximum.

Listings will be provided with flags indicating out of laboratory range values.

5.8.4. Analysis of Vital Signs

Vital signs (including temperature, blood pressure, pulse, and respiration) will be summarized by Baseline and change from Baseline to each scheduled assessment time with descriptive statistics.

5.8.5. Physical Exams

As physical examination is limited in this study, only the targeted examination of lung auscultation will be provided in patient listing.

5.9. Analysis of Pharmacokinetics, Pharmacodynamics and Biomarker data

5.9.1. Analysis of Drug Concentration Data

The concentrations of sarilumab and sIL-6R over time and selected PK parameters, as appropriate, will be summarized using descriptive statistics.

No formal statistical hypothesis testing will be performed.

5.9.2. Analysis of Pharmacodynamic and Exploratory Biomarker data

The concentrations of exploratory PD/Biomarkers over time will be summarized using descriptive statistics.

Pharmacodynamic analysis may include a summary of the time to nadir (or peak) ANC, descriptive statistics of absolute value, absolute change from baseline, percent change from baseline by nominal time (visit), and area under the curve (AUC) concentration of mean and median change from baseline for IL-6 and ANC in serum at each time point. No formal statistical hypothesis testing will be performed.

6. DATA CONVENTIONS

The following analysis conventions will be used in the statistical analysis.

6.1. Definition of Baseline for Efficacy/Safety Variables

Definitions of baseline for efficacy variables are defined in Section 4.5.

For safety variables, baseline will be the latest available valid measurement taken prior to the administration of study drug.

6.2. Data Handling Convention for Efficacy Variables

6.3. Data Handling Convention for Missing Data

Rules for handling missing data for primary and secondary efficacy variables are described in Section 4.5.1, Section 4.5.2 and Section 4.5.3.

For other continuous variables not mentioned in above sections, missing clinical efficacy data will be imputed using last observation carried forward (LOCF) as these are hospitalized patients and missing data is assumed to be missing at random.

For categorical variables, patients with missing data will be included in calculations of percentages. Number of patients with missing data will be presented.

Handling of Medications with missing/partial dates

To determine whether a medication is prior or concomitant medication, the missing medication start date is estimated as early as possible up to first dose date, and the missing medication end date is estimated as late as possible up to Day 29. If the medication start date is missing, the onset day will not be imputed in medication listings.

Handling of Adverse events Severity and Relatedness

If the intensity of a SAE and AESI is missing, it will be classified as “severe” in the frequency tables by intensity of SAE and AESIs. If the assessment of relationship of the investigational product is missing, it will be classified as related to the investigational product.

Date of infusions

Date of infusion is the non-missing administration date filled in the Study Drug Administration-IV CRF. If the first dose of study drug administration date is missing (even after site is queried), then the dosing date is will be imputed with the randomization date. If any subsequent study drug administration date is missing, the date of dispensation of study drug from IRT will be used.

6.4. Visit Windows

Following windows will be used for summarizing laboratory parameters.

Table 5: Time Window for Summary of Laboratory Parameters

Visit label	Target Day	Hematology, Chemistry
Baseline	1	≤1
Day 4	4	2-5
Day 7	7	6-10
Day 15	15	11-17
Day 21	21	18-24
Day 29	29	≥25

6.5. Unscheduled Assessments

The determination of baselines and values at the end of treatment for both efficacy and safety variables will be based on scheduled available assessments and unscheduled available assessments.

Extra assessments (e.g., laboratory data or vital signs associated with non-protocol clinical visits or obtained in the course of investigating or managing adverse events) will be included in listings, but not summaries except for the endpoint determination. If more than one laboratory value is available for a given visit, the first observation will be used in summaries and all observations will be presented in listings.

6.6. Pooling of Centers for Statistical Analyses

Additional analyses on primary and key secondary efficacy endpoints may be performed by sites and pooling sites enrolling less than 5 patients as exploratory analyses.

7. INTERIM ANALYSIS

Because of the unmet medical need for effective medicines in this global COVID-19 pandemic situation, interim analyses are planned for efficacy in the Phase 3 portion of the study. The timing of the first interim analysis will be determined once approximately 110 patients in the Phase 3 Cohort 1 critical stratum requiring mechanical ventilation at baseline and randomized to 400 mg or placebo are enrolled, with a data lock point 22 days later. Interim analysis for this data lock will be conducted in all critical Phase 3 patients who were enrolled until the first dose date of the last 110th patient in Cohort 1 critical stratum randomized to sarilumab 400 mg or placebo who was on a ventilator at baseline.

In the Phase 3 Cohort 1, the primary efficacy endpoint will be tested at interim analysis at the 0.005 (2-sided) significance level. This will correspond to 65% information fraction at the first interim look (ie, 110 patients out of 170 at final). Under the Hwang-Shih-DeCani alpha spending function (Hwang, 1990 [6], Anderson, 2020 [1]), if the interim analysis is not significant, then the significance level for testing the final analysis of primary endpoint will be **0.049** (2-sided). The overall Type 1 error with this alpha spending function is preserved at 0.05 (2-sided). (See Appendix 10.4.2).

Should the interim analysis be positive, the alpha from the interim analysis will be reallocated and added to the alpha for the planned key secondary endpoints which will allow them to be tested at a significance level of 0.05.

Similar interim analyses at approximately 50% to 70% information fraction are planned for the Phase 3 Cohort 2 and Cohort 3 (one interim look per cohort).

8. SOFTWARE

Statistical analyses will be done using SAS Version 9.4. Some analyses may also be done using the R language.

9. REFERENCES

1. Anderson KM. Package "gsDesign" version 3.1.1, May 7, 2020. Available at: http://cran.r-project.org/web/packages/gsDesign/gsDesign.pdf . Accessed: May 13, 2020.
2. ICH. (1996, July 30). ICH Harmonized tripartite guideline: Structure and content of clinical study reports (E3). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.
3. ICH. (1997, July 17). ICH Harmonized tripartite guideline: General considerations for clinical trials (E8). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.
4. ICH. (1998, February 5). ICH Harmonized tripartite guideline: Statistical principles for clinical trials (E9). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.
5. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet. 2020 Feb 15;395(10223):497-506.
6. Hwang IK, Shih WJ, and DeCani JS. Group sequential designs using a family of type 1 error probability spending functions. Statistics in Medicine, 9:1439–1445, 1990.
7. Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. Statistics in Medicine 1990; 9(12): 1447-54.
8. Garner W. Constructing Confidence Intervals for the Differences of Binomial Proportions in SAS®. Gilead Sciences. https://www.lexjansen.com/wuss/2016/127_Final_Paper_PDF.pdf.
9. Kim DH, Uno H, Wei LJ. Restricted mean survival time as a measure to interpret clinical trial results. JAMA cardiology. 2017 Nov 1;2(11):1179-80.
10. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: Consider cytokine storm syndromes and immunosuppression. Lancet 2020. Vol 395 (10229), p1033-1034.
11. Peterson RL, Vock DM, Powers JH, Emery S, Cruz EF, Hunsberger S, et al. Analysis of an ordinal endpoint for use in evaluating treatments for severe influenza requiring hospitalization. Clin Trials 2017; 14(3):264-76.
12. Xu X, Han M, Li T, Sun W, Wang D, Fu B, Zhou Y et al. "Effective treatment of severe COVID-19 patients with Tocilizumab." ChinaXiv preprint (2020). https://www.ser.es/wp-content/uploads/2020/03/TCZ-and-COVID-19.pdf
13. Zhang Huiping. Proportion difference and confidence interval based on Cochran-Mantel-Haenszel method in stratified multi-center clinical trial. PharmSUG China 2016 – Paper 25. https://www.lexjansen.com/pharmasug-cn/2016/PS/PharmaSUG-China-2016-PS07.pdf
14. Zhang W, Zhao Y, Zhang F, et al., The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The experience of clinical immunologists from China, Clinical Immunology (2020), https://doi.org/10.1016/j.clim.2020.108393.

10. APPENDIX

10.1. Schedule of Events

Study Procedure	**Screening Visit[1]**	**Baseline Visit[1]**	**Daily Follow-up Until Hospital Discharge**					**EOS[2]**
Day	**-1 or 1**	**1**	**2**	**4**	**7**	**8 to 23**	**29 and discharge[3,4]**	**60**
Window							±7 days	±7 days
Screening/Baseline:								
Inclusion/Exclusion	X							
Informed Consent	X							
[redacted]	[redacted]							
Demographics and Medical History[1]	X							
Randomization		X						
Treatment:								
Study Drug Administration		X	X[5] If no clinical response			X[5] Repeat weekly for patients requiring supplemental O_2		
Assessments:								
Oxygen administration (FiO_2) and Oxygenation (SpO_2)[6]	X	X	Phase 2 & Phase 3 Cohort 1 - 2 times a day Phase 3 Cohorts 2 and 3 - At least once per day and any time clinical status changes (eg, endotracheal intubation/extubation or requires change in oxygen delivery device)					
Clinical Status Assessment (7-point ordinal scale)[7]		X	Daily until discharge					
NEWS2 Score								
Air or oxygen		X	Daily in the morning until discharge					
Respiratory rate		X	Daily in the morning until discharge					
BP		X	Daily in the morning until discharge					
Pulse		X	Daily in the morning until discharge					
Consciousness		X	Daily in the morning until discharge					
Body temperature before antipyretics or 4 hours after antipyretics[8]		X	Daily in the morning until discharge					
Imaging, microbiology results, and arterial			If available in the medical record					

Study Procedure	**Screening Visit[1]**	**Baseline Visit[1]**	**Daily Follow-up Until Hospital Discharge**					**EOS[2]**
Day	**-1 or 1**	**1**	**2**	**4**	**7**	**8 to 23**	**29 and discharge[3,4]**	**60**
Window							±7 days	±7 days
blood gas results (as available)[9]								
Limited physical examination (lung auscultation only)	X							
Electrocardiogram (ECG), if feasible[10]	X							
Record Targeted Medications[11]	X		Daily until discharge					
Adverse Events[12]	X		X					
Pregnancy Test (WOCBP)[13]	X							
Follow-up Phone Call							X	X
Laboratory Testing:								
C-Reactive Protein (**mandatory**)	X	X		X	X			
			Order CRP with all routine labs and obtain at the same time Record all results when available in medical chart					
Hematology[14]	X		Must be collected within 48 hours prior to redosing Record all results when available in medical chart					
Blood chemistry (including LFTs and creatinine)[15]	X		Must be collected within 48 hours prior to redosing Record all results when available in medical chart					
Ferritin, LDH, d-dimer		X (**mandatory**)	When available					
Blood cultures for bacteria and fungi[16]					X	X *Weekly culture based on ANC[16]*		
PK/Biomarkers/Research (defer to footnotes for sampling requirements):								
	-1 or 1	**1**	**2**	**4**	**7**	**8 to 23 (Only collect on day of dosing)**	**29 or discharge[3,4]**	**60**
Serum for PK/Sarilumab Concentration[17]		X	X	X	X	X	X	
Serum sIL-6R plus research[17]		X		X	X	X	X	
Serum cytokines including IL-6 and biomarker testing[17]		X		X	X	X	X	
Blood for PCR SARS-CoV-2[17,18]	X	X		X			X	

Study Procedure	Screening Visit[1]	Baseline Visit[1]	Daily Follow-up Until Hospital Discharge					EOS[2]
Day	-1 or 1	1	2	4	7	8 to 23	29 and discharge[3,4]	60
Window							±7 days	±7 days
Oropharyngeal or nasopharyngeal swab for SARS-COV-2 detection and sequencing[17,18]	X	X		X			X	
Blood for research plasma[17]		X					X	
[redacted]								
[redacted]		X						

Footnotes for the Schedule of Events Table

1. Screening and baseline may occur on the same day. Assessments that are noted for both visits should only be assessed once. Medical history should include collecting onset of pneumonia symptoms. Body temperature, SpO_2, and FiO_2 must be collected at randomization.
2. Patients will have an end of study (EOS) assessment to collect data on survival and history of hospital re-admission. This assessment may be performed by phone.
3. Patients discharged prior to Day 29 will have a follow-up phone call on Day 29 to collect data on survival and history of hospital re-admission and do not need to be recalled to the hospital for a visit.
4. Patients discharged prior to Day 29 should have a sample collected. If day of discharge is not Day 29 and coincides with another visit, the Day 29 assessments should be performed.
5. Patients will be re-dosed according to Section 6.1 of the protocol.
6. Oxygen administration and oxygenation: refer to Section 9.2.2.3 of the protocol for details. SpO_2 must be measured after 5 minutes of rest (sitting or supine) and must be measured simultaneously with oxygen administration and ventilation data. Record oxygen flow rate (L/min) for patients receiving nasal cannula, simple face mask, or non-rebreather mask. Record FiO_2 for patients receiving high flow nasal cannula, non-invasive ventilation, mechanical ventilation, or extracorporeal membrane oxygenation.
7. Clinical Status Assessment using the 7-point ordinal scale: refer to Section 9.2.2.4 of the protocol for details.

8. Temperature may be measured using the following methods: oral, rectal, tympanic, or temporal according to local hospital protocols and according to the manufacturer's instructions for use of the device. Body temperature should be measured using the same method each time. Temperature should be measured predose after at least 5 minutes of rest (supine or sitting).
9. If available in the medical record, chest CT images will be collected as part of a separate effort related to this study for predictive exploratory analysis and may be provided in a separate study report.
10. ECG only if feasible. Historical ECG from current hospital admission is acceptable.
11. Targeted medications: refer to Section 9.2.3.4 of the protocol for details.
12. Adverse events: Only SAEs and AESIs will be recorded in eCRF.
13. Pregnancy testing to be performed in women of childbearing potential (WOCBP) only. Serum or urine pregnancy test are both acceptable.
14. Hematology: refer to Section 9.2.3.5 of the protocol for details. CBC is required prior to randomization (standard of care labs may be used). After Day 1, CBC will not be performed as a study procedure. When CBC is performed as part of the patient's clinical care, the results will be entered in eCRF.
15. Blood Chemistry: refer to Section 9.2.3.5 of the protocol for details. LFTs and creatinine are required prior to randomization (standard of care labs may be used). After Day 1, LFTs and creatinine will not be performed as a study procedure. When chemistries are performed as part of the patient's clinical care, the results will be entered in eCRF.
16. Surveillance blood cultures for bacteria and fungi should be performed weekly for patients who have had a sustained ANC <1000/μL for >48 hours post-randomization.
17. All samples should be collected before study drug administration except post-infusion PK and sIL-6R samples:

- PK and sIL-6R samples collected on dosing days for the initial dose and for the FIRST repeat dose are mandatory. Samples for subsequent doses are requested if sufficient PPE are available:

— One predose (as close to initiation of treatment as reasonable) and

— One within 60 minutes after the end-of-infusion (EOI). The EOI sample or flush should be collected from the arm, contralateral to that used for IV infusion, if possible. If not medically feasible, the sample can be drawn from the same arm. If the sample cannot be obtained within 60 minutes of the end of infusion, the time from end of infusion should be provided in the CRF

— Day 4 samples are mandatory (if PPE and appropriate lab facilities are available).

- The Day 1 predose sample and Day 29 or Early Termination PK sample may be used for ADA analysis.

18. Swab and tests will be for exploratory analysis only not for inclusion or diagnosis.

[redacted].

10.2. Criteria for Potentially Clinically Significant Values (PCSV)

Parameter	PCSV	Comments
Clinical Chemistry		
ALT*	>3 and ≤ 5 ULN and baseline ≤ 3 ULN* >5 and ≤ 10 ULN and baseline ≤ 5 ULN >10 and ≤ 20 ULN and baseline ≤ 10 ULN >20 ULN and baseline ≤ 20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Each category is calculated independently. * At least one level is required; multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution across the different PCSV levels, additional shift table on ≤3, >3 to ≤5, > 5 to ≤10, >10 to ≤20, and > 20 category for baseline vs. post baseline may be provided
AST*	>3 and ≤ 5 ULN and baseline ≤ 3 ULN* >5 and ≤ 10 ULN and baseline ≤ 5 ULN >10 and ≤ 20 ULN and baseline ≤ 10 ULN >20 ULN and baseline ≤ 20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Each category is calculated independently. * At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution across the different PCSV levels, additional shift table on ≤3, >3 to ≤5, > 5 to ≤10, >10 to ≤20, and > 20 category for baseline vs. post baseline may be provided
Alkaline Phosphatase	>1.5 ULN and baseline ≤ 1.5 ULN	Enzyme activity must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007.

Parameter	PCSV	Comments
Total Bilirubin*	>1.5 and ≤ 2 ULN and baseline ≤ 1.5 ULN* >2 ULN and baseline ≤ 2.0 ULN	Must be expressed in ULN, not in µmol/L or mg/L. Categories are cumulative. Concept paper on DILI – FDA draft Guidance Oct 2007. * At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution of significant level, additional shift table on ≤1.5, >1.5 to ≤2.0 and > 2.0 category for baseline vs. post baseline may be provided
Conjugated Bilirubin	>35% Total Bilirubin and TBILI>1.5 ULN, and baseline Total Bilirubin ≤ 35% or TBILI ≤1.5 ULN	Conjugated bilirubin determined on a case-by-case basis.
ALT and Total Bilirubin	ALT>3 ULN and TBILI>2 ULN, and baseline ALT ≤ 3 ULN or TBILI ≤ 2ULN	Concept paper on DILI – FDA draft Guidance Oct 2007.
CPK*	>3 and ≤ 10 ULN and baseline ≤ 3ULN* >10 ULN and baseline ≤10ULN	FDA Feb 2005. Am J Cardiol April 2006. Categories are cumulative. * At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution of significant level, additional shift table on ≤3, >3 to ≤10, and > 10 category for baseline vs. post baseline may be provided
Creatinine	≥150 µmol/L (Adults) or ≥ULN (if ULN≥150 µmol/L) and baseline < 150 µmol/L or <ULN (if ULN≥150 µmol/L) ≥30% change from baseline ≥100% change from baseline	Benichou C., 1994. 3 independent criteria

Parameter	PCSV	Comments
Creatinine Clearance (Cockcroft's formula)	<15 ml/min and baseline ≥15 ml/min (end stage renal impairment) ≥15 -<30 ml/min and baseline ≥30 ml/min (severe renal impairment) ≥30 - < 60 ml/min and baseline ≥60 ml/min (moderate renal impairment) ≥60 - < 90 ml/min and baseline ≥90 ml/min (mild renal impairment)	Use is optional. FDA draft guidance 2010 Four independent criteria, will provide additional shift table if needed
Uric Acid Hyperuricemia: Hypouricemia:	 >408 µmol/L or >ULN (if ULN≥408 µmol/L) and baseline ≤408 µmol/L or ≤ULN (if ULN≥408 µmol/L) <120 µmol/L or <LLN (if LLN≤120 µmol/L) and baseline ≥ 120 µmol/L or ≥LLN (if LLN≤120 µmol/L)	Harrison- Principles of Internal Medicine 17th Ed., 2008. Two independent criteria
Blood Urea Nitrogen	≥17 mmol/L or ≥ULN (if ULN≥17 mmol/L) and baseline <17 mmol/L or <ULN (if ULN≥17 mmol/L)	Two independent criteria
Chloride Hypochloremia: Hyperchloremia:	 <80 mmol/L or <LLN (if LLN≤80 mmol/L) and baseline ≥ 80 mmol/L or ≥LLN (if LLN≤80 mmol/L) >115 mmol/L or >ULN (if ULN≥115 mmol/L) and baseline ≤ 115 mmol/L or ≤ULN (if ULN≥115 mmol/L)	Two independent criteria
Sodium Hyponatremia: Hypernatremia:	 ≤129 mmol/L or ≤LLN (if LLN≤129 mmol/L) and baseline > 129 mmol/L or >LLN (if LLN≤129 mmol/L) ≥160 mmol/L or ≥ULN (if ULN≥160 mmol/L) and baseline <160 mmol/L or <ULN (if ULN≥160 mmol/L)	Two independent criteria

Parameter	PCSV	Comments
Potassium Hypokalemia Hyperkalemia	 <3 mmol/L or <LLN (if LLN≤3 mmol/L) and baseline ≥ 3 mmol/L or ≥LLN (if LLN≤3 mmol/L) ≥5.5 mmol/L or ≥ULN (if ULN≥5.5 mmol/L) and baseline <5.5 mmol/L or <ULN (if ULN≥5.5 mmol/L)	FDA Feb 2005. Two independent criteria
Total Cholesterol	≥7.74 mmol/L or ≥ULN (if ULN≥7.74 mmol/L) and baseline < 7.74 mmol/L or <ULN (if ULN≥7.74 mmol/L)	Threshold for therapeutic intervention.
Triglycerides	≥4.6 mmol/L or ≥ULN (if ULN≥4.6 mmol/L) and baseline < 4.6 mmol/L or <ULN (if ULN≥4.6 mmol/L)	Threshold for therapeutic intervention.
Lipasemia	≥3 ULN and baseline < 3 ULN	
Amylasemia	≥3 ULN and baseline < 3 ULN	
Glucose Hypoglycaemia Hyperglycaemia	 ≤3.9 mmol/L and <LLN and baseline >3.9 mmol/L or ≥ LLN ≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted) and baseline < 11.1 mmol/L (unfasted); <7 mmol/L (fasted)	ADA Jan 2008.
HbA1c	>8% and baseline ≤ 8%	
Albumin	≤25 g/L or ≤LLN (if LLN≤25 g/L) and baseline >25 g/L or >LLN (if LLN≤25 g/L)	
CRP	>2 ULN or >10 mg/L (if ULN not provided) and baseline ≤2 ULN or ≤10 mg/L (if ULN not provided)	FDA Sept 2005.
Hematology		
WBC	<3.0 Giga/L or <LLN (if LLN≤3.0 Giga/L) and baseline ≥3.0 Giga/L or ≥LLN (if LLN≤3.0 Giga/L) (Non-Black); <2.0 Giga/L or <LLN (if LLN≤2.0 Giga/L) and baseline ≥2.0 Giga/L or ≥LLN (if LLN≤2.0 Giga/L) (Black)* ≥16.0 Giga/L or ≥ULN (if ULN≥16.0 Giga/L) and baseline < 16 Giga/L or <ULN (if ULN≥16.0 Giga/L)	Increase in WBC: not relevant. *The default criteria. Summary by race (black and Non-black) are optional. To be interpreted only if no differential count available.

Parameter	PCSV	Comments
Lymphocytes	>4.0 Giga/L or >ULN (if ULN≥4.0 Giga/L) and baseline ≤ 4.0 Giga/L or ≤ULN (if ULN≥4.0 Giga/L)	
Neutrophils	<1.5 Giga/L or <LLN (if LLN≤1.5 Giga/L) for Non-Black or <1.0 Giga/L or <LLN (if LLN≤1.0 Giga/L) for Black and baseline ≥1.5 Giga/L or ≥LLN (if LLN≤1.5 Giga/L) for Non-Black or ≥1.0 Giga/L or ≥LLN (if LLN≤1.0 Giga/L) for Black* <1.5 Giga/L or <LLN (if LLN≤1.5 Giga/L) and baseline ≥1.5 Giga/L or ≥LLN (if LLN≤1.5 Giga/L) (Non-Black); <1.0 Giga/L or <LLN (if LLN≤1.0 Giga/L) and baseline ≥1.0 Giga/L or ≥LLN (if LLN≤1.0 Giga/L) (Black) <500 Giga/L regardless of baseline value or racce	International Consensus meeting on drug-induced blood cytopenias, 1991. *The default criteria. By race (black and Non-black) are optional.
Monocytes	>0.7 Giga/L or >ULN (if ULN≥0.7 Giga/L) and baseline ≤ 0.7 Giga/L or ≤ULN (if ULN≥0.7 Giga/L)	
Basophils	>0.1 Giga/L or >ULN (if ULN≥0.1 Giga/L) and baseline ≤ 0.1 Giga/L or ≤ULN (if ULN≥0.1 Giga/L)	
Eosinophils	>0.5 Giga/L or >ULN (if ULN≥0.5 Giga/L) and baseline ≤0.5 Giga/L or ≤ULN (if ULN≥0.5 Giga/L)	Harrison- Principles of Internal Medicine 17th Ed., 2008.

Parameter	PCSV	Comments
Hemoglobin	≤115 g/L or ≤LLN (if LLN≤115 g/L) for male or ≤95 g/L or ≤LLN (if LLN≤95 g/L) for female and baseline > 115 g/L or >LLN (if LLN≤115 g/L) for male or > 95 g/L or >LLN (if LLN≤95 g/L) for Female* ≤115 g/L or ≤LLN (if LLN≤115 g/L) and baseline > 115 g/L or >LLN (if LLN≤115 g/L) for male; ≤95 g/L or ≤LLN (if LLN≤95 g/L) and baseline > 95 g/L or >LLN (if LLN≤95 g/L) for Female. ≥185 g/L or ≥ULN (if ULN≥185 g/L) for male or ≥165 g/L or ≥ULN (if ULN≥165 g/L) for female and baseline <185 g/L or <ULN (if ULN≥185 g/L) for male or <165 g/L or <ULN (if ULN≥165 g/L) for Female* ≥185 g/L or ≥ULN (if ULN≥185 g/L) and baseline <185 g/L or <ULN (if ULN≥185 g/L) for Male; ≥165 g/L or ≥ULN (if ULN≥165 g/L) and baseline < 165 g/L or <ULN (if ULN≥165 g/L) for Female Decrease from Baseline ≥20 g/L	Three criteria are independent. *The default criteria. By gender (male and female) are optional. Criteria based upon decrease from baseline are more relevant than based on absolute value. Other categories for decrease from baseline can be used (≥30 g/L, ≥40 g/L, ≥50 g/L).

Parameter	PCSV	Comments
Hematocrit	≤0.37 v/v or ≤LLN (if LLN≤0.37 v/v) for Male or ≤0.32 v/v or ≤LLN (if LLN≤0.32 v/v) for Female and baseline > 0.37 v/v or >LLN (if LLN≤0.37 v/v) for Male or > 0.32 v/v or >LLN (if LLN≤0.32 v/v) for Female* ≤0.37 v/v or ≤LLN (if LLN≤0.37 v/v) and baseline > 0.37 v/v or >LLN (if LLN≤0.37 v/v) for Male ; ≤0.32 v/v or ≤LLN (if LLN≤0.32 v/v) and baseline > 0.32 v/v or >LLN (if LLN≤0.32v/v) for Female ≥0.55 v/v or ≥ULN (if ULN≥0.55 v/v) for Male or ≥0.5 v/v or ≥ULN (if ULN≥0.5 v/v) for Female and baseline < 0.55 v/v or <ULN (if ULN≥0.55 v/v) for Male < 0.5 v/v or <ULN (if ULN≥0.5 v/v) for Female* ≥0.55 v/v or ≥ULN (if ULN≥0.55 v/v) and baseline < 0.55 v/v or <ULN (if ULN≥0.55 v/v) for Male ; ≥0.5 v/v or ≥ULN (if ULN≥0.5 v/v) and baseline < 0.5 v/v or <ULN (if ULN≥0.5 v/v) for Female	Two Criteria are independent *The default criteria. By gender (male and female) are optional.
RBC	≥6 Tera/L or ≥ULN (if ULN≥6 Tera/L) and baseline < 6 Tera/L or <ULN (if ULN≥6 Tera/L)	Unless specifically required for particular drug development, the analysis is redundant with that of Hb.
Platelets	<100 Giga/L or <LLN (if LLN≤100 Giga/L) and baseline ≥100 Giga/L or ≥LLN (if LLN≤100 Giga/L) ≥700 Giga/L or ≥ULN (if ULN≥700 Giga/L) and baseline < 700 Giga/L or <ULN (if ULN≥700 Giga/L)	International Consensus meeting on drug-induced blood cytopenias, 1991. Two independent criteria
Urinalysis		
pH	≤4.6 or ≤LLN (if LLN≤4.6) and baseline > 4.6 or >LLN (if LLN≤4.6) ≥8 or ≥ULN (if ULN≥8) and baseline < 8 or <ULN (if ULN≥8)	Two independent criteria
Vital signs		
HR	<45 bpm and decrease from baseline ≥20 bpm ≥120 bpm and increase from baseline≥20 bpm	To be applied for all positions except STANDING

Parameter	PCSV	Comments
SBP	≤95 mmHg and decrease from baseline ≥20mmHg ≥160 mmHg and increase from baseline ≥20 mmHg	To be applied for all positions except STANDING
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg ≥110 mmHg and increase from baseline ≥10 mmHg	To be applied for all positions except STANDING
Weight	≥ 5% increase from baseline ≥5% decrease from baseline	FDA Feb 2007

10.3. Criteria for Relevant Major Protocol Deviations

The relevant major protocol deviations leading to exclusion of patients from the Per Protocol Set (PPS) are listed below.

Table 6: Protocol Deviations List

Category	Description of Protocol Deviation	Protocol Deviation Identification (PDID)
Entered study even though entry criteria was not satisfied	Incl1: Patient was not >=18 years of age at time of enrollment but was randomized	1.01
Entered study even though entry criteria was not satisfied	Incl2a: Patient did not have evidence of pneumonia by chest radiograph, chest computed tomography or chest auscultation (rales, crackles), requires supplemental oxygen and/or ventilation	1.02
Entered study even though entry criteria was not satisfied	Incl2b: Patient did not meet at least one of the criterion of Severe disease, Critical disease, Multi-organ dysfunction, or Immunocompromised category at Baseline Visit	1.03
Entered study even though entry criteria was not satisfied	Incl3. Patient did not have laboratory confirmed SARS-CoV-2 infection as determined by PCR, result from any specimen (or other commercial or public health assay) within 2 weeks prior to randomization	1.04
Entered study even though entry criteria was not satisfied	Incl4. Patient did not comply with clinic visits and study - related procedures	1.05
Entered study even though entry criteria was not satisfied	Incl5. Patient or legally acceptable representative did not provide informed consent	1.06
Entered study even though entry criteria was not satisfied	Excl1. Patient, in the opinion of the investigator, not expected to survive for >= 48 hours from screening but was randomized	1.07
Entered study even though entry criteria was not satisfied	Excl2a. Absolute neutrophil count (ANC) less than 2000mm3	1.08
Entered study even though entry criteria was not satisfied	Excl2b. AST or ALT great than 5 x ULN	1.09
Entered study even though entry criteria was not satisfied	Excl2c. Platelets <50,000 per mm3	1.10
Entered study even though entry criteria was not satisfied	Excl3. Treatment with anti-IL 6, anti IL 6R antagonists, or with a Janus kinase inhibitors (JAKi) in the past 30 days or plans to receive during the study period	1.11
Entered study even though entry criteria was not satisfied	Excl4. Current treatment with the simultaneous combination of leflunomide and methotrexate	1.14
Entered study even though entry criteria was not satisfied	Excl5a. Known active tuberculosis or a history of incompletely treated tuberculosis	1.17
Entered study even though entry criteria was not satisfied	Excl5b. Suspected or known extrapulmonary tuberculosis	1.18
Entered study even though entry criteria was not satisfied	Excl6. Patients with suspected or known active systemic bacterial or fungal infections	1.19

Category	Description of Protocol Deviation	Protocol Deviation Identification (PDID)
Entered study even though entry criteria was not satisfied	Excl8. Participation in any clinical research study within 3 months and < 5 half-lives of investigational product prior to the Screening Visit	1.21
Entered study even though entry criteria was not satisfied	Excl9. Any physical exam findings and/or history of any illness that, might confound the results of the study or pose risk to the patient by participating the study	1.22
Entered study even though entry criteria was not satisfied	Excl10. Known systemic hypersensitivity to sarilumab or the excipients of the drug product	1.23
Subject developed withdrawal criteria but were not withdrawn	Patient experienced AEs listed in protocol section 8.3.1.2 that warrants permanent termination of the IV study drug infusions, but continued to receive study drug	2.01
Received wrong treatment or incorrect dose	Subject was given study drug but was not randomized	4.01
Received wrong treatment or incorrect dose	Subject received incorrect treatment (e.g., wrong kit given, kit dispensed without IWRS transaction)	4.02
Received wrong treatment or incorrect dose	Subject received incorrect dose administration or unacceptable IP (expired or temperature excursion deemed unacceptable)	4.03
Other Treatment compliance	Staff inappropriately unblinded to treatment assignment	5.01
Other Treatment compliance	Treatment assignment inappropriately unblinded	5.02
Randomization Error	Stratification error - not stratified by severity of illness at enrollment or systemic corticosteroids for COVID-19	6.01
Visit not performed	Baseline Visit not performed	7.02
Inadequate Informed Consent administration	Patient or legally acceptable representative did not sign an ICF and study procedures initiated (never signed ICF or signed after procedure)	11.01
Inadequate Informed Consent administration	Patient or legally acceptable representative did not sign an amended ICF and study procedures initiated	11.02
Inadequate Informed Consent administration	Patient confidentiality not maintained	11.03
Inadequate Informed Consent administration	Study Procedures performed after patient withdrew full consent	11.04

10.4. Program codes (SAS or R codes)

10.4.1. SAS code for primary efficacy variables

The following code is written in the SAS programming language for the statistical analysis of proportions comparing two groups. Statistical method for testing the two proportions is the stratified Cochran-Mantel-Haenszel test with SAS code shown below.

For calculating the confidence intervals for differences in two proportion using the strata-adjusted CMH method, refer to the SAS code in Zhang, 2016 [13].

```
data sampledata;
  input strata trtgrp $ outcome $ count @@;
  datalines;
1 placebo 1Yes 18 1 placebo 2No 26
2 placebo 1Yes 12 2 placebo 2No 13
1 drug1   1Yes 40 1 drug1   2No 54
2 drug1   1Yes 26 2 drug1   2No 26
1 drug2   1Yes 51 1 drug2   2No 37
2 drug2   1Yes 35 2 drug2   2No 25
;
run;

/* Analysis using "strata" as stratification factor */
proc freq data=sampledata(where=(trtgrp in ('placebo','drug2')));
  weight count;
  tables strata*trtgrp*outcome / cmh nopercent nocol;
run;
```

10.4.2. R code for alpha spending at interim analysis

The following code is written in the **R** programming language and uses the R package **gsDesign** [1] which is used to design group sequential clinical trials.

```
install.packages("gsDesign")
library(gsDesign)

#Assume 2 looks: 1 Interim Look and 1 Final Look
#Data will be available for 110 patients at Interim Analysis, out
of 170 at Final Analysis
#Information fraction for binomial endpoint is 65% (=110/170).

#Fix nominal alpha=0.005 at interim
x.Phase3.IA.option1 <- gsDesign(k=2, test.type=4, alpha=0.025,
beta=0.1, timing=c(110/170,1), sfupar=-6.5)
x.Phase3.IA.option1
```

The output of the above R code is shown in Figure 3. As displayed in the output, the nominal significance level at 1st interim look will be α=2*0.0025=0.005 (2-sided) (ie, corresponds to critical boundary for efficacy (upper bound), Z = 2.8070). At the final look, the significance level will be α = 2*0.0245 = 0.049 (2-sided) (ie, corresponds to critical Z=±1.9660 for a 2-sided test).

Figure 3: Sample Output for a Group Sequential Design with Interim Analysis

```
> #Fix nominal alpha=0.005 at interim
> x.Phase3.IA.option1 <- gsDesign(k=2, test.type = 4, alpha=0.025, beta=0.1, timing=c(110/170,1), sfupar=-6.5)
> x.Phase3.IA.option1
Asymmetric two-sided group sequential design with
90 % power and 2.5 % Type I Error.
Upper bound spending computations assume
trial continues if lower bound is crossed.

          Sample
           Size    ----Lower bounds----  ----Upper bounds-----
  Analysis Ratio*  Z   Nominal p Spend+  Z   Nominal p Spend++
         1  0.678 0.93    0.8251 0.0414 2.81    0.0025  0.0025
         2  1.048 1.97    0.9755 0.0586 1.97    0.0245  0.0225
     Total                       0.1000                0.0250
+ lower bound beta spending (under H1):
 Hwang-Shih-DeCani spending function with gamma = -2.
++ alpha spending:
 Hwang-Shih-DeCani spending function with gamma = -6.5.
* Sample size ratio compared to fixed design with no interim

Boundary crossing probabilities and expected sample size
assume any cross stops the trial

Upper boundary (power or Type I Error)
          Analysis
   Theta      1      2  Total   E{N}
  0.0000 0.0025 0.0208 0.0233 0.7418
  3.2415 0.4445 0.4555 0.9000 0.8681

Lower boundary (futility or Type II Error)
          Analysis
   Theta      1      2  Total
  0.0000 0.8251 0.1516 0.9767
  3.2415 0.0414 0.0586 0.1000
```

Signature Page for VV-RIM-00110347 v1.0

ESig Approval	[redacted]-2020 16:13:44 GMT+0000
ESig Approval	[redacted]-2020 16:17:19 GMT+0000
ESig Approval	[redacted]-2020 16:20:54 GMT+0000
ESig Approval	[redacted]000
ESig Approval	[redacted]000
ESig Approval	[redacted]000

Signature Page for VV-RIM-00110347 v1.0 Approved

Signature Page for VV-RIM-00110347 v1.0

ESig Approval	[redacted]000

ESig Approval	[redacted]000

Signature Page for VV-RIM-00110347 v1.0 Approved