

Official Title: A Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate the Safety and Efficacy of Tocilizumab in Patients With Severe COVID-19 Pneumonia

NCT Number: NCT04320615

Document Date: SAP Version 3: 16-July-2020

STATISTICAL ANALYSIS PLAN

TITLE: A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY TO EVALUATE THE SAFETY AND EFFICACY OF TOCILIZUMAB IN PATIENTS WITH SEVERE COVID 19 PNEUMONIA

PROTOCOL NUMBER: WA42380

STUDY DRUG: Tocilizumab (RO4877533)

VERSION NUMBER: 3

IND NUMBER: 148225

EUDRACT NUMBER: 2020-001154-22

SPONSOR: F. Hoffmann-La Roche Ltd.

PLAN PREPARED BY: [REDACTED]

DATE FINAL: Version 1: 24 April 2020

DATES AMENDED Version 2: 26 May 2020
Version 3: See electronic date stamp below

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Date and Time(UTC)	Reason for Signing	Name
16-Jul-2020 16:46:46	Company Signatory	[REDACTED]

CONFIDENTIAL

This is an F. Hoffmann-La Roche Ltd document that contains confidential information. Nothing herein is to be disclosed without written consent from F. Hoffmann-La Roche Ltd

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

This Statistical Analysis Plan WA42380 Version 3 was amended from Version 2 as follows:

- Time to event endpoints were changed from “time from randomization” to “time from first dose of study drug”
- The Type I error control section was updated to specify a hierarchy for testing of the primary endpoint followed by testing the difference in mortality
- Cumulative Incidence Function plots were specified for time to ‘improvement’ end points
- The censoring rules for the time to event endpoints were updated
- Derivation of organ failure-free days was clarified
- Laboratory ranges were clarified
- ‘Other’ was removed from the stratification by region as only patients from Europe and North America were randomized.
- The synopsis appended from the protocol was updated to be consistent with protocol version 3.
- A subgroup analysis of the primary endpoint by mechanical ventilation status (as stratified) was added to the primary analysis section

Additional minor changes have been made to improve clarity and consistency.

This Statistical Analysis Plan WA42380 Version 2 was amended from Version 1 as follows:

- The sample size was amended to 450 patients based on powering the study at 90%
- A new secondary endpoint was added ‘Time to Recovery’
- Time to discharge or “ready for discharge” was elevated to one of the key secondary endpoints
- The censoring rules for deaths were changed to right censoring for time to event endpoints, other than for time to clinical failure
- The derivation of vent-free days was modified so that patients that have died are assigned zero vent-free days.
- The list of Adverse Events of Special Interest was modified

Additional minor changes have been made to improve clarity and consistency.

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE.....	2
1. BACKGROUND	8
2. STUDY DESIGN	8
2.1 Protocol Synopsis.....	10
2.2 Endpoints.....	10
2.2.1 Primary Efficacy Endpoints.....	10
2.2.2 Secondary Efficacy Endpoints.....	10
2.2.2.1 Assessments Specific to National Early Warning Score 2	11
2.2.3 Exploratory Efficacy Endpoints	11
2.2.4 Pharmacodynamic Efficacy Endpoints.....	11
2.2.5 Biomarkers	11
2.2.6 Safety Endpoints	12
2.3 Determination of Sample Size	12
2.4 Analysis Timing	13
3. STUDY CONDUCT	13
3.1 Randomization, Stratification and Blinding	14
3.2 Data Monitoring	15
4. STATISTICAL METHODS	15
4.1 Analysis Populations	16
4.1.1 mITT Population	16
4.1.2 Safety Population	16
4.2 Analysis of Study Conduct.....	16
4.3 Analysis of Treatment Group Comparability	16
4.3.1 Demographics	17
4.3.2 Disease Characteristics	17
4.3.3 Medical history.....	18
4.3.4 Surgeries and Procedures	18
4.3.5 Previous and Concomitant Medications.....	18
4.4 Visit Labels	18
4.5 Efficacy Analysis.....	20

4.5.1	Primary Efficacy Endpoint.....	21
4.5.2	Controlling for Type I Error	22
4.5.3	Secondary Endpoints	23
4.5.3.1	Time to Event Analyses	23
4.5.3.2	Incidence endpoints.....	27
4.5.3.3	Duration endpoints	28
4.5.4	Exploratory Efficacy Endpoints	30
4.5.5	Subgroup Analyses	30
4.6	Pharmacodynamic Analyses	31
4.7	Safety Analyses	31
4.7.1	Exposure of Study Medication.....	32
4.7.2	Adverse Events	32
4.7.3	Laboratory Data.....	33
4.7.4	Vital Signs.....	34
4.7.5	Other Safety Endpoints.....	34
4.8	Interim Analyses	35
5.	REFERENCES	37

LIST OF TABLES

Table 1	Distribution of Ordinal Scale in the Placebo Group	13
Table 2	Distribution of the Ordinal scale in the Tocilizumab Group	13
Table 3	Time Windows for Assigning Assessment Study Days to Study Visits Labels for Deaths and Discharge	19
Table 4	Time Windows for Assigning Assessment Study Days to Study Visits for PD parameters	20
Table 5	Time to clinical improvement and Censoring	24
Table 6	Time to improvement in Ordinal Clinical Status and Censoring ..	25
Table 7	Time to Recovery and censoring	25
Table 8	Time to hospital discharge or “ready for discharge” and censoring.	26
Table 9	Time to Clinical Failure and censoring	27

LIST OF FIGURES

Figure 1	Study WA42380 Schema	9
----------	----------------------------	---

LIST OF APPENDICES

Appendix 1	Protocol Synopsis	38
Appendix 2	Schedule of Activities: Days 1 and 2	45
Appendix 3	Schedule of Activities: Days 3–28	49
Appendix 4	Schedule of Activities: After Day 28	52
Appendix 5	National Early Warning Score 2 (NEWS2)	54
Appendix 6	Cochran-Mantel-Haenszel Test	56

GLOSSARY OF ABBREVIATIONS

AE	adverse event
AEGT	adverse event grouped term
AESI	adverse event of special interest
BAL	bronchoalveolar lavage
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
COVID-19	coronavirus 2019/ SARS-CoV-2
CRP	C-reactive protein
DMC	data monitoring committee
ECG	electrocardiogram
ECMO	extracorporeal membrane oxygenation
eCRF	electronic Case Report Form
ICU	intensive care unit
IL-6	interleukin 6
iDMC	independent data monitoring committee
ISAP	interim statistical analysis plan
IxRS	interactive voice or web-based response system
LPLV	last patient last visit
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent to treat
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NEWS2	National Early Warning Score 2
PaO ₂ /FiO ₂	Ratio between the blood pressure of the oxygen (partial pressure of oxygen, PaO ₂) and the percentage of oxygen supplied (fraction of inspired oxygen, FiO ₂)
PBO	placebo
PCR	polymerase chain reaction
PD	pharmacodynamic
PK	pharmacokinetic
PT	preferred term
RT-PCR	reverse-transcriptase polymerase chain reaction
SAE	serious adverse event
SAP	statistical analysis plan
sIL-6R	soluble interleukin-6 receptor
SMQ	Standard MedDRA Query

SMT study management team
SOC standard of care / system organ class
SoC scientific oversight committee
SpO2 blood oxygen saturation
TB tuberculosis
TCZ tocilizumab
TLR top line report
TTCI time to clinical improvement
VFDs ventilator-free days
WHO World Health Organisation

1. BACKGROUND

This Statistical Analysis Plan (SAP) provides details of the planned analyses and statistical methods for the clinical efficacy and clinical safety for Study WA42380. Any analyses of biomarkers will be covered by a separate analysis plan. Analyses of pharmacokinetic data will be covered by a separate analysis plan.

There are currently no drugs licensed for the treatment of patients with SARS-CoV-2 (COVID-19). Based on the results from an initial 21-patient retrospective observational study, in which patients with severe or critical COVID-19 pneumonia were treated with tocilizumab (TCZ) off-label (Xu et al. 2020), TCZ, along with standard of care (SOC) treatment, could provide efficacy, offering the potential benefit to treat COVID-19 in hospitalized populations; with the limitation for this observational study of a lack of a proper control as a comparator. Extensive safety data have previously been generated on the use of TCZ in other indications. Therefore, a randomized placebo-controlled study in combination with SOC to assess safety and efficacy of TCZ in hospitalized patients with severe COVID-19 pneumonia is justified to address the high unmet need and burden of disease in this severely ill population.

2. STUDY DESIGN

Study WA42380 is a Phase III, randomized, double-blind, placebo-controlled, multicenter study to assess the efficacy and safety of TCZ in combination with SOC compared with matching placebo (PBO) in combination with SOC in hospitalized adult patients with severe COVID-19 pneumonia. The Sponsor intends to enroll approximately 450 patients that diagnosed with COVID-19 pneumonia that meet the entry criteria in centers globally.

Patients must be at least 18 years of age with confirmed SARS-CoV-2 (COVID-19) infection per WHO criteria, including a positive polymerase chain reaction (PCR) of any specimen (e.g., respiratory, blood, urine, stool, other bodily fluid). At the time of enrollment, patients must have blood oxygen saturation (SpO_2) $\leq 93\%$ or PaO_2/FiO_2 (the ratio between the blood pressure of the oxygen [partial pressure of oxygen, PaO_2] and the percentage of oxygen supplied [fraction of inspired oxygen, FiO_2] < 300 mmHg despite being on SOC, which may include anti-viral treatment, low dose steroids, and supportive care.

Patients in whom, in the opinion of the treating physician, progression to death is imminent and inevitable within the next 24 hours, irrespective of the provision of treatments, will be excluded from the study. Patients with active tuberculosis (TB) or suspected active bacterial, fungal, viral, or other infection (besides COVID-19) will be excluded from the study.

Patients will be randomized at a 2:1 ratio to receive blinded treatment of either TCZ or a matching placebo, respectively. Study treatment will be given in combination with SOC per local practice. The randomization will be stratified by geographic region (North

America, Europe) and mechanical ventilation (yes, no). The proportion of patients on a mechanical ventilator will be capped at no more than 50% of the overall study population.

Patients assigned to the TCZ arm will receive one infusion of TCZ 8 mg/kg, with a maximum dose of 800 mg, and patients assigned to the placebo arm will receive one infusion of placebo, both in addition to SOC.

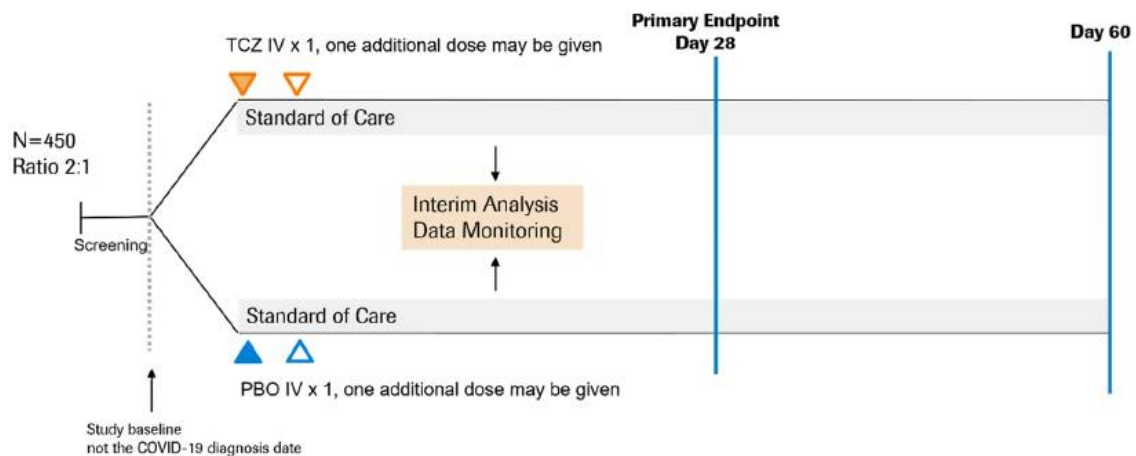
For both arms, if the clinical signs or symptoms worsen or do not improve (reflected by sustained fever or at least a one-category worsening on the 7-category ordinal scale), one additional infusion of blinded treatment of TCZ or placebo can be given, 8–24 hours after the initial infusion.

Patients who do not meet the criteria for participation in this study (screen failure) may qualify for one re-screening opportunity (for a total of two screenings per participant) at the investigator's discretion. The investigator will record the reasons for screen failure in the screening log.

The study assessments to be conducted include the following: physical examination, vital signs, oxygen saturation, assessment of consciousness, presence and absence of respiratory support, adverse events, concomitant therapies, clinical laboratory tests, and nasopharyngeal swabs. Please see [Appendix 2](#), [Appendix 3](#) and [Appendix 4](#) for details concerning the timing of these assessments.

[Figure 1](#) presents an overview of the study design. The Schedule of Assessments is provided in [Appendix 2](#), [Appendix 3](#) and [Appendix 4](#).

Figure 1 Study WA42380 Schema



IV = intravenous; PBO = placebo; TCZ = tocilizumab.

2.1 PROTOCOL SYNOPSIS

The Protocol Synopsis is in [Appendix 1](#). For additional details, see the Schedule of Assessments in [Appendix 2](#), [Appendix 3](#) and [Appendix 4](#).

2.2 ENDPOINTS

This study will evaluate the efficacy, safety, pharmacodynamics, and pharmacokinetics of TCZ compared with a matching placebo in combination with SOC in hospitalized patients with severe COVID-19 pneumonia. Specific objectives and corresponding endpoints for the study are outlined below.

2.2.1 Primary Efficacy Endpoints

The primary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

- Clinical status assessed using a 7-category ordinal scale at Day 28

The ordinal scale categories are as follows:

1. Discharged (or “ready for discharge” as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or \leq 2L supplemental oxygen)
2. Non-intensive care unit (ICU) hospital ward (or “ready for hospital ward”) not requiring supplemental oxygen
3. Non-ICU hospital ward (or “ready for hospital ward”) requiring supplemental oxygen
4. ICU or non-ICU hospital ward, requiring non-invasive ventilation or high-flow oxygen
5. ICU, requiring intubation and mechanical ventilation
6. ICU, requiring extracorporeal membrane oxygenation (ECMO) or mechanical ventilation and additional organ support (e.g. vasopressors, renal replacement therapy)
7. Death

2.2.2 Secondary Efficacy Endpoints

- Time to clinical improvement (TTCI) defined as a National Early Warning Score 2 (NEWS2) of \leq 2 maintained for 24 hours
- Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical status
- Incidence of mechanical ventilation
- Ventilator-free days to Day 28
- Incidence of intensive care unit (ICU) stay

- Duration of ICU stay
- Clinical status assessed using a 7-category ordinal scale at Day 14
- Time to clinical failure, defined as the time to death, mechanical ventilation, ICU admission, or withdrawal (whichever occurs first). For patients entering the study already in ICU or on mechanical ventilation, clinical failure is defined as a one-category worsening on the ordinal scale, withdrawal or death.
- Mortality at Days 7, 14, 21, 28, and 60
- Time to hospital discharge or “ready for discharge” (as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or $\leq 2\text{L}$ supplemental oxygen)
- Time to recovery defined as hospital discharge or “ready for discharge” (as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or $\leq 2\text{L}$ supplemental oxygen), or Non-ICU hospital ward (or “ready for hospital ward”) not requiring supplemental oxygen
- Duration of supplemental oxygen

2.2.2.1 Assessments Specific to National Early Warning Score 2

In addition to the vital measurements, the patient’s consciousness level and the presence or absence of respiratory support must be recorded. The NEWS2 parameter for respiratory support is the selection of either air or “oxygen”, which can include other forms of ventilation to maintain oxygen saturation (see [Appendix 5](#)).

NEWS2 values will be calculated by the Sponsor based on vital sign parameters and NEWS2 related assessments recorded by the investigator in the appropriate electronic Case Report Form (eCRF).

2.2.3 Exploratory Efficacy Endpoints

- Incidence of vasopressor use
- Duration of vasopressor use
- Incidence of extracorporeal membrane oxygenation (ECMO)
- Duration of ECMO
- Organ failure-free days

2.2.4 Pharmacodynamic Efficacy Endpoints

- Serum concentrations of interleukin 6 (IL-6), soluble interleukin-6 receptor (sIL-6R), ferritin, and C-reactive protein (CRP) at specified time points as shown in the schedule of assessments (see [Appendix 2](#), [Appendix 3](#) and [Appendix 4](#)).

2.2.5 Biomarkers

The exploratory biomarker objectives for this study are to identify and/or evaluate biomarkers that could be predictive of response to TCZ (i.e., predictive biomarkers), may serve as early surrogates of efficacy, may be associated with progression to a more

severe disease state (i.e., prognostic biomarkers), may be associated with susceptibility to developing AEs or could lead to improved adverse event (AE) monitoring or investigation (i.e., safety biomarkers), could further evidence of TCZ pharmacological activity (i.e., pharmacodynamic biomarkers), and overall increase our knowledge and understanding of disease pathogenesis and drug safety, on the basis of the following endpoint:

- Exploratory analysis of individual biomarkers in relation to efficacy, safety, exposure (listed in Section “Laboratory, Biomarker, and Other Biological Samples” of the protocol) and in both blood- and tissue-derived samples will be defined in a separate SAP.

2.2.6 Safety Endpoints

- Incidence and severity of AEs, with severity determined according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) v5.0
- COVID-19 (SARS-CoV-2) viral load over time, as collected by nasopharyngeal swab and bronchoalveolar lavage (BAL) samples (if applicable)
- Time to reverse-transcriptase polymerase chain reaction (RT-PCR) virus negativity
- The proportion of patients with any post-treatment infection
- Change from baseline in targeted clinical laboratory test results

2.3 DETERMINATION OF SAMPLE SIZE

The estimated sample size was determined for the primary endpoint of comparison of clinical status based on a 7-category ordinal scale at Week 4 using the Van Elteren test. [Table 1](#) shows the assumed distribution of the ordinal scale in the PBO plus SOC group. [Table 2](#) shows the expected distribution in the TCZ plus SOC group with an odds ratio of 2 (assuming proportional odds). Under these assumptions, the total modified intent to treat (mITT) sample size of 450 with a 2:1 randomization of TCZ to placebo patients provides approximately 90% power to detect a difference in distribution between the treatment groups of the ordinal scale at Week 4 using a two-sided Van Elteren test at the 5% significance level.

In addition this sample size provides approximately 90% power to detect a ratio of 2 (TCZ to PBO) for the odds of being in a category or a better category under the assumptions of the expected probability distribution of patients in the placebo arm in [Table 1](#), using a proportional odds model with a two-sided p-value at the 5% significance level.

Assuming proportional odds and the given distribution of the placebo group, the smallest odds ratio that could be statistically significant would be approximately 1.5.

This sample size also provides approximately 90% power to detect a 10% absolute difference in mortality rate under the assumption of a 15% mortality rate in the placebo group.

Table 1 Distribution of Ordinal Scale in the Placebo Group

1 (discharge)	2	3	4	5	6	7 (death)
0.58	0.05	0.09	0.09	0.02	0.02	0.15

Assuming proportional odds the expected distribution in the TCZ arm with an odds ratio of 2 would be:

Table 2 Distribution of the Ordinal scale in the Tocilizumab Group

1 (discharge)	2	3	4	5	6	7 (death)
0.734	0.039	0.064	0.058	0.012	0.012	0.081

2.4 ANALYSIS TIMING

Up to three interim looks for efficacy prior to the primary analysis will be carried out on the data with mortality rate at 4 Weeks (secondary endpoint) evaluated for interim efficacy analyses. The interim looks will occur after roughly 111, 222 and 333 patients have been enrolled and have reached the Day 28 follow-up time point, but all interims are subject to change depending on the enrollment rate. For additional information about interim analyses, refer to Section 4.8.

If efficacy is declared based on an interim analysis of mortality, the data will be cleaned, a snapshot taken and the data in the snapshot will be reported. There will then be a final snapshot when all patients either reach Day 60, or have withdrawn.

If the study does not meet the efficacy criteria at one of the interim looks, or the efficacy interim is not performed, no reports for interim data, other than for the data monitoring committee (DMC), will be prepared; a snapshot of the data will be taken and the primary analysis will occur when the last patient either has withdrawn or completed the Day 28 visit. A clinical study report (CSR) and/or a top line report (TLR) based on the analyses from this snapshot will be produced.

There will be an additional analysis on the final data when all patients have either reached Day 60 or withdrawn. Analyses from the first reporting event, restricted to data up to Day 28 (Week 4), will not be updated based on the final snapshot.

3. STUDY CONDUCT

The plan is to enroll approximately 450 patients that have been diagnosed with COVID-19 pneumonia and meet the entry criteria in centers globally. Patients will be

randomized at a 2:1 ratio to receive blinded treatment of either TCZ or a matching placebo, respectively. Study treatment must be given in combination with standard of care. For both arms, if the clinical signs or symptoms worsen or do not improve, one additional infusion of blinded treatment of TCZ or placebo can be given, 8–24 hours after the initial infusion.

Patients who do not meet the criteria for participation in this study (screen failure) may qualify for one re-screening opportunity (for a total of two screenings per participant) at the investigator's discretion.

Patients will be followed up for a total of 60 days after first dose of study medication.

If patients are discharged from hospital prior to Day 28, follow up visits should occur twice weekly through Day 28. Patients are encouraged to return for onsite visits at least weekly if possible; other follow-up visits may be conducted by telephone or home visits. Patients should return to the site for a Day 28 visit. After Day 28, all patients should have follow up visits on Day 35, Day 45, and Day 60; the Day 35 and Day 45 visits may be conducted by telephone or by home visits for discharged patients, while the Day 60 visit should be conducted onsite.

During the study, standard supportive care will be given according to clinical practice.

3.1 RANDOMIZATION, STRATIFICATION AND BLINDING

Patients will be randomized as soon as possible after screening at a 2:1 ratio to receive blinded treatment of either TCZ or a matching placebo, respectively. Study treatment must be given in combination with SOC. The randomization will be stratified by geographic region (North America, Europe) and mechanical ventilation (yes, no); and will occur through use of a permuted-block randomization method to ensure a balanced assignment to each treatment arm. The proportion of patients who are on a mechanical ventilator at the time of randomization will be capped at no more than 50% of the overall study population.

Study site personnel and patients will be blinded to treatment assignment during the study. The Sponsor and its agents will also be blinded to treatment assignment, with the exception of individuals who require access to patient treatment assignments to fulfill their job roles during a clinical trial. These roles include the unblinding group responsible, clinical supply chain managers, sample handling staff, operational assay group personnel, interactive voice or web-based response system (IxRS) service provider, and Data Monitoring Committee (DMC) members and statistical programming analysts working with the DMC.

While pharmacokinetic (PK) samples must be collected from patients assigned to the comparator arm to maintain the blinding of treatment assignment, PK results for these patients are generally not needed for the safe conduct or proper interpretation of the

study data. Laboratories responsible for performing study drug PK assays will be unblinded to patient treatment assignments to identify appropriate samples for analysis. PK samples from patients assigned to the comparator arm will not be analyzed for study drug PK concentration except by request (e.g., to evaluate a possible error in dosing).

If unblinding is necessary for a medical emergency (e.g., in the case of a serious adverse event [SAE] for which patient management might be affected by knowledge of treatment assignment), the investigator will be able to break the treatment code by contacting the IxRS. The investigator is not required to contact the Medical Monitor prior to breaking the treatment code; however, the treatment code should not be broken except in emergency situations.

As per health authority reporting requirements, the Sponsor's Drug Safety representative will break the treatment code for all serious, unexpected suspected adverse reactions that are considered by the investigator or Sponsor to be related to study drug. The patient may continue to receive treatment, and the investigator, patient, and Sponsor personnel, with the exception of the Drug Safety representative and personnel who must have access to patient treatment assignments to fulfill their roles (as defined above), will remain blinded to treatment assignment.

3.2 DATA MONITORING

A DMC will monitor the incidence of all SAEs, adverse events of special interest (AESI) and any anticipated events during the study.

The DMC will also evaluate safety according to policies and procedures detailed in a DMC Charter. Regular safety reviews will begin after approximately 15 patients (10 TCZ, 5 placebo) have been enrolled and reached 14-day follow-up. Early stopping criteria will be detailed in the DMC charter and a separate interim statistical analysis plan (ISAP). Further details of efficacy interims are provided in Section 4.8. Interim analyses will be conducted by a statistical programmer independent from the study management team and will be reviewed by the DMC. Interactions between the DMC and Sponsor will be carried out as specified in the DMC Charter.

The Data Monitoring Committee will initially consist of Sponsor representatives not directly involved in the study management team (SMT) and a scientific oversight committee (SoC) of external experts (responsibilities and operating principles of the DMC are described in a charter, the Internal Monitoring Committee and Scientific Oversight Committee Agreement). If feasible during study conduct the DMC responsibilities may transition to a fully independent data monitoring committee (iDMC).

4. STATISTICAL METHODS

All primary and secondary efficacy endpoints will be analyzed in the mITT population, with patients grouped according to the treatment assignment at randomization.

In all safety and pharmacodynamic analyses patients will be grouped according to the treatment that the patients actually received rather than the treatment assigned at randomization.

4.1 ANALYSIS POPULATIONS

Disposition summaries will be based on an All Patient population (all patients randomized and/or receiving study drug). Efficacy analyses will be based on the mITT population, if not otherwise specified. Analysis of safety data and pharmacodynamic (PD) data will be based on the safety population.

4.1.1 mITT Population

The mITT population is defined as all patients randomized in the study that received any amount of study medication, with patients grouped according to the treatment assignment at randomization.

4.1.2 Safety Population

Safety population will consist of all patients who receive any amount of study medication. In all safety and pharmacodynamic analyses, patients will be grouped according to the treatment that the patients first received rather than the treatment assigned at randomization.

4.2 ANALYSIS OF STUDY CONDUCT

The number of patients enrolled, discontinued, or who complete the study will be summarized to week 4 and to the end of the study. Reasons for premature study discontinuation will be listed and summarized to Week 4; and additionally to the end of the study. Listing of randomized patients and a listing of investigators will be produced.

The number of patients discharged from hospital will also be summarized by visit.

Eligibility criteria and other major protocol deviations will be listed and summarized by treatment group.

A listing by treatment group and patient of missed assessments for the primary endpoint will be produced through to Day 28, including study day of missed assessment, study day of discharge and /or death, if any.

The patients excluded from the safety and mITT populations will be summarized, including the reason for exclusion by treatment group. A summary of enrollment by country and investigator name will be produced.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic and baseline characteristics (including, but not limited to, age, sex, race, geographic region, NEWS2, ordinal scale for clinical status, IL-6, sIL-6R, mechanical ventilation, anti-viral treatment at baseline, steroids at baseline) will be summarized

using means, standard deviations, medians, and ranges for continuous variables and proportions for categorical variables, as appropriate. Summaries will be presented by treatment group and will be presented for the mITT and may, in addition, be presented for the safety population.

4.3.1 Demographics

- Sex
- Age
- Weight
- Race
- Ethnicity
- Geographic region
- Female fertility status

4.3.2 Disease Characteristics

- Smoking history (Never, Current, Former)
 - Former/current user: number of years subject smoked (years), Nicotine exposure in pack years
 - e-cigarettes use (Yes/No)
- NEWS2
- Ordinal scale for clinical status
- IL-6
- CRP
- Ferritin
- sIL-6R
- Mechanical ventilation
- Steroid use at Day 1 (to be derived from concomitant medication)
- Anti-viral treatment at Day 1 (to be derived from concomitant medication)
- Symptoms at time of COVID 19 diagnosis
 - Fever
 - Cough
 - Shortness of breath
 - gastrointestinal symptoms (e.g. diarrhea, nausea, loss of appetite)
 - Headache
 - Fatigue
 - Other
- Number of days from first COVID-19 symptom at baseline

- COVID-19 diagnosis based on PCR of specimen type
- Number of days from COVID-19 diagnosis at baseline
- Specimen type at screening
- PCR result (Negative, positive)
- Quantitative PCR result (viral load)

4.3.3 Medical history

Medical history data will be summarized descriptively by treatment group using the safety population. A glossary showing the mapping of investigator verbatim terms to diseases will be produced for the medical history data.

4.3.4 Surgeries and Procedures

A listing of any previous or ongoing surgeries and procedures will be produced for the safety population.

4.3.5 Previous and Concomitant Medications

Previous and concomitant treatments will be summarized descriptively by treatment group for the safety population. Previous treatments that have been stopped prior to study Day 1 will be summarized separately. There will be a summary of all concomitant treatments, including those that were initiated prior to study day 1. In addition there will be a summary of all treatments with the indication given as 'COVID-19'.

Previous and concomitant treatments will be listed, with treatments for COVID-19 listed separately.

A summary of patients requiring supplemental oxygen post-discharge will be provided. This will be based on "home oxygen" being recorded on the Concomitant Medication page of the eCRF.

A glossary showing the mapping of investigator verbatim terms to medication coded terms will be produced for previous or concomitant medication.

4.4 VISIT LABELS

For summaries of data not collected by visit, such as AEs, medical history and concomitant medications all data up to the end of study will be included. Exceptions to this include death, discharge and ICU stay; which will be summarized weekly in descriptive summaries, following the time windowing approach described below.

Deaths will also be captured on the ordinal scale of clinical status. Deaths confirmed by public record are also captured in the eCRF, which may not have been captured as AEs for patients withdrawn from the study. These events will also be incorporated into the windowing for death.

Table 3 Time Windows for Assigning Assessment Study Days to Study Visits Labels for Deaths and Discharge

Scheduled study day	Efficacy time window
1* (Baseline)	< 1
7 (Week 1)	1 to 7
14 (Week 2)	>7 to ≤ 14
21 (Week 3)	> 14 to ≤ 21
28 (Week 4)	> 21 to ≤ 28
35 (Week 5)	> 28 to ≤ 35
45	> 35 to ≤ 45
60	> 45 to ≤ 67

*Study day 1 is the first day of study drug

Patient assessments that are collected at scheduled visits will be assigned to a study visit using the actual study day of the assessment; this includes data from withdrawal visits and any unscheduled visits. Time windows will be continuous from the midpoint between two consecutive study visits to the next midpoint, and will be dependent on the schedule of assessments for each variable independently. An example of time windowing for the PD parameters (CRP, IL-6, sIL-6R) is shown below.

Table 4 Time Windows for Assigning Assessment Study Days to Study Visits for PD parameters

Scheduled study day	^a Efficacy time window
1 (Baseline)	≤ 1
Day 2	2
Day 3	3
7 (Week 1)	> 3 to ≤ 10
14 (Week 2)	> 10 to ≤ 17
21 (Week 3)	> 17 to ≤ 24
28 (Week 4)	> 24 to ≤ 28
35 (Week 5)	> 28 to ≤ 38
Day 45	> 38 to ≤ 52
Day 60	> 52 to ≤ 67

a From Week 1 onwards use value nearest to scheduled study day.

Where there is more than one efficacy assessment within a time window, then the nearest non-missing assessment will be assigned to that visit. If two or more assessments are equidistant from the scheduled time point, then the latest assessment will be used for efficacy (other than death or discharge where the assessment prior to the visit week will be used as described previously).

For safety parameters such as laboratory parameters and vital signs the ‘worst case’ will be used.

The last value from screening will be used for baseline assessments if there is no baseline (study Day 1) value. Pretreatment assessments will be used preferentially on study Day 1 for baseline.

4.5 EFFICACY ANALYSIS

All efficacy analyses will use the mITT population.

Sensitivity analyses to evaluate the robustness of results to the primary analysis methods (e.g., handling of withdrawals) may be conducted and are described in this SAP in each relevant section.

Descriptive subgroup analyses to evaluate the consistency of results across pre-specified subgroups may also be conducted as specified in Section [4.5.5](#).

4.5.1 Primary Efficacy Endpoint

The primary efficacy objective for this study is to evaluate the efficacy of TCZ plus SOC, compared with placebo plus SOC using the following endpoint:

- Clinical status assessed using a 7-category ordinal scale at Week 4

Assessment of patient status using this ordinal scale will be recorded at baseline and once daily in the morning (between 8 am and 12 pm) while hospitalized.

The primary estimand attributes are:

- Population: Patients with severe COVID-19 pneumonia as per the inclusion/exclusion criteria specified in the protocol (mITT)
- Primary endpoint: Clinical status at Week 4
- Treatments: TCZ plus SOC versus Placebo plus SOC
- Intercurrent events: Events leading to study withdrawal
- Handling of intercurrent events: last observed post-baseline value (except if the patient has been discharged [without re-admittance] or has died up to and including Day 28, then the death or discharge will override the Week 4 value or be imputed for a missing Week 4 value).
- Summary measure: medians (95% CI) PBO plus SOC and TCZ plus SOC

Intercurrent events are those that occur after treatment initiation and either preclude observation of the variable or affect its interpretation. For patients who withdraw before Week 4, their last post baseline ordinal category prior to withdrawal will be used in the primary analysis, unless death within the time frame was captured from public records or otherwise; in which case death will be used in the analysis.

The estimand is the difference in distributions between Tocilizumab plus SOC and Placebo plus SOC which will be tested using a non-parametric method, the Van Elteren test, including the stratification factors at randomization (region [North America, Europe] and mechanical ventilation [yes, no]). The median ordinal scale result for each treatment group and the corresponding 95% CI for the median will be presented along with the Van Elteren P-value, as well as the difference in medians and a 95% CI for the difference.

Additionally, for patients that withdraw prior to Week 4 on the TCZ arm, the 50th percentile of the TCZ data will be imputed from those that complete the study to Week 4 (deaths and discharges are included as completing to Week 4), and on the placebo arm the 50th percentile will be imputed from the placebo data.

For patients that withdraw prior to Week 4 on the TCZ arm the 75th percentile will be imputed from those that complete the study to Week 4 (deaths and discharges are included as completing to Week 4), and on the placebo arm the 25th percentile will be imputed from the placebo data.

In addition, the clinical status according to the 7-category ordinal scale will be compared between the TCZ group and the placebo group at Week 4, using a proportional odds model accounting for stratification factors at randomization in the model (region [North America, Europe] and mechanical ventilation [yes, no] using the mITT population. The odds ratio, p-value, and 95% confidence interval will be presented.

The assumption of proportional odds will be evaluated by visually comparing the fitted proportions of patients across the ordinal scale from the model with the observed data.

In addition to imputing the ordinal scale at Week 4 with an earlier death or discharge (without re-admittance), captured from ordinal scale or other sources, this imputation rule will also be followed at earlier time points, including the day of death or discharge.

A death or discharge (unless the patient is re-admitted) will always be carried forward to all subsequent assessments regardless of what is recorded for the ordinal scale. If a patient is re-admitted then the ordinal scale data from the point of re-admittance will be used. The ordinal scale will be summarized by Week and treatment showing n and percentage in each category, as well as missing data. Comparison of clinical status according to the 7-category ordinal scale (detailed for the primary endpoint at Week 4) will be analyzed using a proportional odds model at additional time points including but not limited to Week 2.

Stacked bar charts of the ordinal scale will be produced by treatment group, the bars will total to 100% and the categories, including 'missing', will be shown. At Week 4, a side by side comparison of the treatment groups by stacked bar chart will be shown.

The primary endpoint will be analysed by the mechanical ventilation status as randomized, as it is considered that this stratification factor may be predictive for response to tocilizumab treatment. The clinical status assessed using a 7-category ordinal scale at Week 4 will be tested in the subgroups split by mechanical ventilation status (yes, no) as stratified using the Van Elteren test for each subgroup. The subgroups will also be analysed using the proportional odds model as specified for the primary endpoint.

4.5.2 Controlling for Type I Error

The following are key secondary endpoints for the study:

- Difference in Mortality at Week 4
- Ventilator-free days at Week 4
- Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical status to Week 4
- Time to hospital discharge or “ready for discharge”

- Clinical status assessed using a 7-category ordinal scale at Week 2

The primary endpoint (the difference in distributions in clinical status between Tocilizumab plus SOC and Placebo plus SOC at Week 4; Van Elteren test) will be tested at a two-sided 5% significance level. If the primary endpoint is statistically significant the difference in mortality at Week 4 will then be tested at 0.05 (two-sided Cochran-Mantel-Haenszel test). There will be no further multiplicity adjustment for the additional four key secondary endpoints.

A treatment effect may be observed in a key secondary endpoint that may not meet nominal statistical significance, but may still be considered clinically meaningful. Therefore all five of the key secondary endpoints will be tested in addition to the primary endpoint in order to help inform prescribers by potentially providing this information in the label if clinically meaningful.

4.5.3 Secondary Endpoints

4.5.3.1 Time to Event Analyses

Time to event secondary endpoints will be compared between the TCZ group and the placebo group using the stratified log-rank test with geographic region (North America, Europe) and mechanical ventilation (yes, no) included as the stratification factors at Day 28 using the mITT population. The Kaplan-Meier plot, median time to response, and their 95% CIs, and a p-value will be presented. In addition, the treatment groups will be compared descriptively using a Cox proportional hazards model adjusting for the stratification factor applied at randomization. Hazard ratios and a 95% CI will be produced.

For time to event endpoints, other than time to clinical failure, deaths will be right censored (at Day 28). Consequently, for these endpoints, participants censored on Day 28 reflect two different states, death and failure to meet the improvement outcome criterion. Therefore, it is important to understand the efficacy outcome in the context of the number and timing of deaths by treatment arm. Cumulative incidence function plots for both death and the event of interest will be produced using the non-parametric Aalen–Johansen estimator.

For time to event endpoints that include discharge as an event, the earliest time of discharge or “ready for discharge” from the different sources of discharge will be used in all analyses. If a patient is discharged and re-admitted more than 12 hours later, then the first discharge will be considered as meeting the event. If a patient is readmitted within 12 hours of discharge, then they will not have met the endpoint at this time. If they are discharged later in the study (without re-admittance within 12 hours), then the later time of discharge will be used.

Time to event endpoints include:

- Time to clinical improvement in hours

Defined as time from first dose of study drug to NEWS2 of ≤ 2 maintained for 24 hours

The estimand is the difference in distributions between Tocilizumab plus SOC and Placebo plus SOC using the log rank test as described above.

Intercurrent events are those that occur after treatment initiation and either preclude observation of the variable or affect its interpretation. Intercurrent events, such as patients who are lost to follow-up, die or discontinue for any reason prior to achieving the event or do not have the event, will be accounted for through censoring rules, as described in [Table 5](#) below.

Patients who have a score of ≤ 2 at baseline will be analysed in the same way as patients with a score that is >2 at baseline.

Partial date times may be imputed based on available data, following a conservative approach. The NEWS2 is to be assessed twice daily, with approximately 12 hours between each assessment. At least two assessments with a score of ≤ 2 covering a span of at least 21.5 hours will be required to meet the criterion, with a maximum of 26.5 hours between the first and last of these assessments (there must be no assessments with a score >2 in between). If a patient has a score of ≤ 2 and is then discharged from hospital within 26.5 hours, with no subsequent scores > 2 before the discharge they will have met the endpoint.

Table 5 Time to clinical improvement and Censoring

Event	Censor	Date and Time
Hospital discharge prior to clinical improvement criterion met	Yes	Hospital discharge
Hospital discharge and hospital re-admission within 12 hours and continue study	No	Not Applicable
Death prior to clinical improvement criterion met	Yes	Day 28
Discontinuation or lost to follow-up for any reason prior to clinical improvement criterion met	Yes	last scheduled vital sign assessment
No clinical improvement	Yes	last vital sign assessment within Week 4 time window

Other time to event endpoints include:

- Time to improvement in ordinal clinical status (days)

Defined as time from first dose of study drug to the time when at least a 2-category improvement in the 7-category ordinal scale is observed. For patients in category 2 at baseline, discharge or “ready for discharge” will be considered as meeting the threshold. For patients that are discharged and the ordinal scale assessment has not been completed at discharge, they will be assumed to be in category one of the ordinal scale at the point of discharge, unless they are re-admitted within 12 hours. Intercurrent events, such as patients who are lost to follow-up, die or discontinue for any reason prior to achieving the event or do not have the event, will be accounted for through censoring rules, as described in [Table 6](#) below.

Table 6 Time to improvement in Ordinal Clinical Status and Censoring

Event	Censor	Date and Time
Hospital discharge and hospital re-admission within 12 hours and continue study	No	Not Applicable
Death prior to improvement in Ordinal clinical status criterion met	Yes	Day 28
Discontinuation or lost to follow up for any reason prior to improvement in Ordinal clinical status criterion met	Yes	last Ordinal scale assessment
No improvement in Ordinal clinical status	Yes	last Ordinal scale assessment within Week 4 time window

- Time from first dose of study drug to Recovery (days)

Defined as discharged or ready for discharge {normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or \leq 2L supplemental oxygen} (ordinal scale category one), or Non-ICU hospital ward {or “ready for hospital ward”} not requiring supplemental oxygen (ordinal scale category 2). Patients that are in category 2 at baseline will need to achieve category 1 to meet the endpoint of recovery.

Intercurrent events, such as patients who are lost to follow-up, die or discontinue for any reason prior to achieving the event or do not have the event, will be accounted for through censoring rules, as described in [Table 7](#) below.

Table 7 Time to Recovery and censoring

Event	Censor	Date and Time
Hospital discharge and hospital re-admission within 12 hours and continue study	No	Not Applicable
Death prior to Recovery	Yes	Day 28
Discontinuation or lost to follow up for any reason prior to recovery criterion met	Yes	last Ordinal scale assessment
Not recovered	Yes	last Ordinal scale assessment within Week 4 time window

- Time from first dose of study drug to hospital discharge or “ready for discharge” (days)

Ready for discharge; defined as normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or \leq 2L supplemental oxygen (ordinal scale category one)

Intercurrent events, such as patients who are lost to follow-up, die or discontinue for any reason prior to achieving the event or do not have the event, will be accounted for through censoring rules, as described in [Table 8](#) below.

Table 8 Time to hospital discharge or “ready for discharge” and censoring

Event	Censor	Date and Time
Hospital discharge and hospital re-admission within 12 hours and continue study	No	Not Applicable
Death prior to discharge	Yes	Day 28
Discontinuation or lost to follow up for any reason prior to discharge or “ready for discharge” criterion met	Yes	last Ordinal scale assessment
Not (discharged or “ ready for discharge”)	Yes	last Ordinal scale assessment within Week 4 time window

- Time to clinical failure (days)

Defined as the time from first dose of study drug to first occurrence on study of death, mechanical ventilation (as collected in the Vital Signs & Oxygen Saturation eCRF), ICU admission or withdrawal (discontinuation from study) prior to discharge, whichever occurs first. For patients entering the study already in ICU or on mechanical ventilation (as collected in the Vital Signs & Oxygen Saturation eCRF), clinical failure is defined as a one-category worsening on the ordinal scale, withdrawal or death. Withdrawals not during hospitalization will not be considered as having met clinical failure and data will be censored at the last vital sign assessment, unless a later death is recorded in which case they will have met the event upon death.

Intercurrent events, such as patients who are lost to follow-up or discontinue for any reason prior to the event or do not have the event, will be accounted for through censoring rules, as described in [Table 9](#) below.

Table 9 Time to Clinical Failure and censoring

Event	Censor	Date and Time
Hospital discharge not followed by death or readmittance	Yes	Last scheduled vital sign assessment (or discharge if no post-discharge vital sign data is available)
Hospital discharge and hospital re-admission within 12 hours and continue study	No	Not Applicable
Lost to follow-up prior to clinical failure criterion met not followed by death	Yes	Last scheduled vital sign assessment or ordinal scale assessment
Clinical failure criterion not met	Yes	last of scheduled vital sign assessments or ordinal scale assessments within Week 4 time window

The NEWS2 score and clinical failure status as defined above will be summarized descriptively by visit.

Sensitivity analyses may be performed considering death as a competitive event (other than for the clinical failure endpoint), using a competing risk model if there is an imbalance between groups in deaths and reason attributed.

4.5.3.2 Incidence endpoints

Secondary efficacy incidence endpoints will be analyzed using the Cochran-Mantel-Haenszel test statistic adjusted by the stratification factors at baseline geographic region (North America, Europe) and mechanical ventilation (yes, no) using the mITT population, unless stated otherwise. The weighted difference in proportions for the treatment group comparison will be presented with a p-value, together with a 95% CI (see [Appendix 6](#)).

- Incidence of mechanical ventilation (as collected in the Vital Signs & Oxygen Saturation eCRF) by Week 4 (mechanical ventilation refers to invasive mechanical ventilation only, and/or ECMO).
- Incidence of ICU stay by Week 4

For incidence of mechanical ventilation or incidence of intensive care stay by Week 4 for patients that have withdrawn or died prior to Week 4, the non-responder rule will be applied, i.e. it will be assumed that the patient required mechanical ventilation, or has had an ICU stay by Week 4 in the analysis. Patients without either mechanical ventilation or intensive care stay respectively prior to discharge, will be assumed to be responders in the analysis, unless the patient is readmitted to hospital within 12 hours, or the patient dies by Week 4.

In addition to the analyses based on the mITT, incidence of ICU stay and mechanical ventilation will also be analyzed excluding those patients that were in ICU/mechanically ventilated (according to the stratification), respectively, at baseline.

The number and proportion of patients requiring mechanical ventilation or an ICU stay will be summarized descriptively by study week.

- Difference in mortality at Week 4

The difference in proportion of patients that have died by Day 28 will be compared using the CMH test as described above. All deaths post discontinuation and discharge will be included in this analysis.

Deaths occurring between each visit, and cumulative deaths by visit will be summarized descriptively to Day 60.

4.5.3.3 Duration endpoints

- Ventilator-free days

The number of Ventilator-free days (VFDs) is defined as the number of days from Day 1 to Day 28 when the patient is alive and breathes without invasive assistance of the mechanical ventilator. VFDs will be derived from the vital signs and oxygen saturation log; if invasive mechanical ventilation or ECMO is recorded for any part of the day, the day will not be counted as a VFD.

VFDs will be zero if the patient is mechanically ventilated from Day 1 to Day 28. VFDs will be zero if a patient dies on or prior to Day 28.

For patients withdrawn early from the study but not discharged, if patients were on invasive-mechanical ventilation at the point of discontinuation it will be assumed that the remainder of days to Week 4 were not VFDs. For patients not using invasive mechanical ventilation at point of withdrawal it will be assumed the period from withdrawal to Week 4 are VFDs. For patients that are discharged, days from discharge to Day 28 will be counted as VFDs. If ventilator data is missing for patients that have not withdrawn, died or discharged, then the last observation post-baseline will be carried forward until the next observation.

VFDs will be analysed using the Van Elteren test, including the stratification factors at randomization (region [North America, Europe] and mechanical ventilation [yes, no]). The median VFDs for each treatment group and the corresponding 95% CI for the median will be presented along with the Van Elteren P-value, as well as the difference in medians and a 95% CI for the difference. A cumulative distribution plot of VFDs will be produced.

VFDs will also be summarized descriptively using the medians, along with 95% CIs, by treatment group for those patients alive at Day 28, with a count of the number of patients assigned zero VFDs due to death by Day 28

- Duration of supplemental O2 (days)

Duration of supplemental O2 (days) will also be derived from the vital signs and oxygen saturation log, where study days with 'supplemental oxygen or other forms of ventilation' will be summed up to and including Week 4. Patients without any supplemental O2 use will assigned a duration of zero days. For missing data, the last observation post-baseline will be carried forward until either the next observation or the point of withdrawal/discharge. For patients withdrawn early from the study but not discharged, if the patients were on supplemental oxygen at the point of withdrawal it will be assumed that the remainder of days to Week 4 were on supplemental oxygen. For patients not using supplemental oxygen at point of withdrawal it will be assumed supplemental oxygen is not required to Week 4. For patients that are discharged, days from discharge to Day 28 will be counted as days without supplemental oxygen (unless supplemental oxygen use is recorded on the Concomitant Medications eCRF during follow up visits, in which case all days from the day after discharge to the end date from the Concomitant Medication eCRF will be classed as days with supplemental oxygen). Duration of supplemental oxygen use will be 28 days if a patient dies on or prior to Day 28.

Days of supplemental O2 use will be analyzed and summarized descriptively in a similar method to VFDs.

- Duration of ICU stay (days)

Duration of ICU stay (days) will be calculated as the sum of the number of hours spent in ICU up to and including Week 4 divided by 24, based on the admission and discharge date times from the ICU stay information log; $(\text{ICU discharge datetime} - \text{ICU admission datetime})/24$. Multiple periods of ICU stay will be summed. Patients without any ICU stays will be assigned a duration of zero days.

Partial admission and discharge times may be imputed based on available data, following a conservative approach. For patients that are discharged, any ongoing ICU stays without an end date will be imputed from date of discharge as appropriate and it will be assumed that days from discharge to Day 28 do not involve an ICU stay. For patients not in the ICU at the point of withdrawal from study it will be assumed that the period to Week 4 has no incidences of ICU stay post withdrawal. For patients in ICU on the day of withdrawal it will be assumed that they are in the ICU throughout the period to Week 4. Patients that die on or prior to Day 28 will be assigned a duration from the first dose of study drug to Day 28 23:59:59.

Days of ICU stay will be analyzed and summarized descriptively in a similar method to VFDs.

4.5.4 Exploratory Efficacy Endpoints

Incidence of vasopressor use (from concomitant medication records) and incidence of extracorporeal membrane oxygenation (ECMO) by Week 4 (and separately to Day 60) will be summarized descriptively.

Duration of vasopressor use (days) and ECMO (days) to Week 4 will be summarized using the median along with 95 % CIs for the median by treatment group. ECMO use is collected daily and the number of days of ECMO use will be totaled. Vasopressor duration will use start and stop dates from the concomitant medication records. A concomitant medication record that is ongoing at Week 4 will use the upper bound of the Week 4 time window as the end date for the duration.

Days without organ failure will be summarized descriptively through Week 4. In addition, a summary of individual organ failure over time will be provided.

Organ failure is defined as present on any day when the most abnormal vital signs/abnormal lab value meets the definition of clinically significant organ failure ([Bernard et al, 1995](#); [NHLBI ARDS Clinical Trials Network 2014](#)). Cardiovascular organ failure is defined as either systolic BP ≤ 90 mmHg or the need for vasopressor. Renal, hepatic and coagulation parameters will be assessed via blood tests in order that the presence of clinically significant organ failure can be determined. Renal failure is defined as creatinine ≥ 2 mg/dl, hepatic failure is defined as bilirubin ≥ 2 mg/dl and coagulation failure is defined as a platelet count of $\leq 80 \times 10^3/\text{mm}^3$. Each day a patient is alive and free of a given clinically significant organ failure will be scored as a failure-free day for that organ. In the case of no data for a particular organ, the last observation post-baseline will be carried forward until the next observation or discharge. Any day that a patient is alive and free of all 4 organ failures (cardiovascular, renal, hepatic, coagulation) will be considered an organ failure-free day.

If a patient dies on or before Day 28, they will be assigned a value of zero organ failure-free days in the overall summary of organ failure-free days. For patients that are discharged, days from discharge to Day 28 will be counted as organ failure-free days, unless they are readmitted in which case the available data will be used.

4.5.5 Subgroup Analyses

A subgroup analysis of the Clinical status assessed using a 7-category ordinal scale at Week 4 by mechanical ventilation status (yes, no) as stratified will be performed using the Van Elteren test for each subgroup. In addition, a stacked barchart at Day 28, a summary of clinical status at Day 28 and a summary of clinical status over time will be produced by mechanical ventilation status (yes, no) at baseline.

In addition, the clinical status according to the 7-category ordinal scale will be compared between the TCZ group and the placebo group at Week 4 by mechanical ventilation status (yes, no) as stratified, using a proportional odds model accounting for stratification

factors at randomization in the model (region [North America, Europe]) using the mITT population. The odds ratio, p-value, and 95% confidence interval will be presented.

The odds ratio for mortality for TCZ versus PBO at day 28 will be analyzed by logistic regression, including covariates of interest as well as the stratification factors in the model. The odds ratio for the treatment effect by gender, race, age (18-64 years, 65-84 years, 85 years and over), region, mechanical ventilation (as stratified) and ordinal clinical status (categories 1 to 3, vs categories 4 and 5) will be determined. Other subgroup analyses may also be performed. If necessary, where there are small subgroups, categories may be collapsed as appropriate to enable analysis, such as for subgroup analysis by race.

4.6 PHARMACODYNAMIC ANALYSES

The PD analysis population will be identical to the safety population.

Summary tables for serum concentration of IL-6, sIL-6R, CRP and ferritin (mean, standard deviation, median, minimum, and maximum) will be produced by visit/ time point and treatment arm.

Individual patient data and descriptive statistics (i.e. median and interquartile range) will be plotted by visit/ time point for each treatment arm.

4.7 SAFETY ANALYSES

Safety assessments will be performed on the safety population. In all safety analyses, patients will be grouped according to the treatment that the patients first received rather than the treatment assigned at randomization.

Descriptive summaries of laboratory values and change from baseline throughout the study will be tabulated by treatment arm. For selected parameters, changes from the proportion of patients experiencing clinically significant changes relative to baseline will be summarized by treatment arm.

Values, along with change from baseline, will be summarized using descriptive statistics for each vital sign parameter.

COVID-19 (SARS-CoV-2) viral load over time, as collected by nasopharyngeal swab and BAL samples (if applicable) will be summarized descriptively by time point and treatment group. The number and proportion of patients negative and positive will be displayed, and for those positive the quantitative result will be summarized.

Time to reverse-transcriptase polymerase chain reaction (RT-PCR) COVID-19 virus negativity will be analyzed using similar methods to the other time to event analyses. There will be an additional analysis limited to patients positive at baseline based on central laboratory results.

4.7.1 Exposure of Study Medication

Exposure to study drug will be summarized including number of patients with one or two doses and number of patients with dose modification by treatment group.

A listing of patients by treatment group will be prepared detailing dosing of study drug, volume administered and any dose modification. If after unblinding a patient has received study drug from more than one treatment group, then the actual dose of TCZ would be calculated from the proportion of TCZ kits received. For example, if a patient received 1 kit of TCZ and 3 kits of PBO then the actual dose of TCZ would be calculated as $\frac{1}{4}$ of the total recorded dose.

4.7.2 Adverse Events

Medical Dictionary for Regulatory Activities (MedDRA) will be used as the thesaurus for AEs and disease codes, and the WHO Drug Global B3 Format dictionary will be used for treatments. A glossary of these codes will be produced.

Only treatment-emergent AEs will be summarized. Treatment-emergent events are defined as those AEs with observed or imputed onset date on or after the start date of trial treatment. Only where the most extreme intensity is greater than the initial intensity (or if most extreme intensity is not missing and initial intensity is missing) will events with an onset date prior to the start of trial treatment be considered treatment-emergent. An AE with a completely missing start date will be assumed to be treatment-emergent unless the AE has a complete non-imputed end date that is prior to study Day 1.

Adverse events will be coded and tabulated by system organ class (SOC), and/or preferred term (PT) and treatment arm. In tabulations, PTs and their associated SOC will be presented in order of descending frequency summed across the treatment arms.

Adverse events will also be tabulated by severity, as graded according to NCI CTCAE v5.0 scale, and relationship to study medication as indicated by the investigator.

The following will also be summarized:

- serious adverse events
- adverse events leading to withdrawal of study drug
- adverse events leading to discontinuation from the study
- adverse events leading to death
- hypersensitivity adverse events (adverse events occurring during or within 24 hours of the end of an infusion that are deemed “related” to study treatment)

Adverse events of special interest will be defined using SOC, published Standard MedDRA Queries (SMQs) or AE Grouped Terms (AEGTs) defined by Roche Drug Safety. The groupings of AEs will include but may not be limited to the following:

- Infections (Infections and Infestations SOC)
- Opportunistic infections (Roche Standard AEGT Basket)
- Malignancies (Malignant or Unspecified tumors SMQ Narrow)
- Hepatic events (Hepatic failure, Fibrosis, and Cirrhosis and Other Liver Damage-related Conditions SMQ Wide or Hepatitis, non-infectious SMQ Wide)
- Stroke (Ischemic Cerebrovascular Conditions SMQ Wide or Hemorrhagic Cerebrovascular SMQ Wide)
- Myocardial infarction [MI] (MI SMQ Wide)
- Anaphylactic reaction events (utilizing Roche Standard AEGT Basket according to Sampson's criteria) [[Sampson et al. 2006](#)] occurring during or within 24 hours of the end of tocilizumab infusion; and a separate summary using the Anaphylactic Reaction SMQ Narrow for events occurring during or within 24 hours of the end of tocilizumab infusion)
- Gastrointestinal perforations (Gastrointestinal perforation SMQ Wide)
- Bleeding events (Hemorrhages SMQ Wide)
- Demyelinating events (Demyelination SMQ Narrow)

Summaries and listings of malignancies and gastrointestinal perforations that were confirmed by medical review will be provided.

A glossary showing the mapping of investigator verbatim terms to preferred terms will be produced for all AEs included in the analysis. For each AE of special interest table based on SMQs/AEGTs, a corresponding listing of the preferred terms that comprise the SMQ will be produced.

Listings of AEs and SAEs will be produced. Adverse events of special interest will also be listed.

AEs and SAEs will be summarized by age category (18-64 years, 65-84 years, 85 years and over).

The exposure duration on study (exposure duration is the date of the last safety assessment or death if present, minus the date of the first dose of TCZ plus one divided by 365.25) will be summarized.

4.7.3 Laboratory Data

Laboratory data will use ranges from local laboratories and laboratory values will be converted to Système International units.

Summary tables will detail the actual values and changes from baseline of the laboratory parameters over visits by treatment arm. Arterial blood gases will be summarized separately. Summaries of the number of patients by CTC grade for hematology and hepatic lab parameters (Alkaline phosphatase [ALP], Alanine aminotransferase [ALT] and aspartate aminotransferase [AST]), and total bilirubin) will be produced (for summaries referring to NCI CTCAE grading).

For neutrophils, platelets, lymphocytes, and hepatic lab parameters the number of patients will be summarized by CTCAE grade category for baseline and worst post baseline result.

Patients with values outside the reference will be listed, with an indication of the direction of the abnormality (High, Low).

A listing of all pregnancies will be presented.

4.7.4 Vital Signs

Summary statistics on absolute values and their change from baseline for all observed vital signs (diastolic blood pressures, systolic blood pressures, respiratory rate, pulse rate, body temperature and peripheral oxygen saturation) will be presented over time by treatment group. Baseline is defined as the last assessment prior to treatment. Additionally, a graphical representation of means over time of oxygen saturation and temperature (daily to Week 4) will be presented.

For patients requiring supplemental oxygen, summary statistics on absolute values of the oxygen flow rate (L/min) and/or fraction of inspired oxygen (FiO₂) will be produced by visit/ time point and treatment group.

The level of consciousness will be summarized over time.

The number and proportion of patients requiring oxygen supplementation or other form of ventilation will be summarized over time, including type of support given. Non-invasive mechanical ventilation will be summarized overall as well as by its component types (continuous positive airway pressure [CPAP], bi-level positive airway pressure [BiPAP], other). Invasive mechanical ventilation will also be summarized overall and by component types (Endotracheal tube, tracheostomy tube).

A listing of patients with chest X-ray, CT scans and ECGs (as a separate listing) with clinically significant abnormalities will be produced.

4.7.5 Other Safety Endpoints

SARS-CoV-2 viral load over time, as collected by nasopharyngeal swab and BAL samples (if applicable) will be summarized descriptively by time point and treatment group.

Time to reverse-transcriptase polymerase chain reaction (RT-PCR) COVID-19 virus negativity will be analyzed using similar methods to the other time to analyses.

4.8 INTERIM ANALYSES

Up to three efficacy interim analyses will be carried out on the data with mortality rate at 28 days (secondary endpoint) evaluated for efficacy. The interim looks will occur after roughly 111, 222 and 333, patients are enrolled and have reached 28 days follow up, but all interims are subject to change depending on the enrollment rate.

The first efficacy interim analysis will be conducted when approximately 111 patients (74 TCZ and 37 placebo) have reached the 28-day follow-up time point and will be based on the mortality rate at 28 days (secondary endpoint). Up to two further interim efficacy analyses will occur after roughly 222 and 333 patients have reached the 28-day follow-up time point. If the results of one of the interim analyses meets the pre-specified criteria for efficacy, further enrollment in the placebo arm will be discontinued and all enrolled patients will be eligible receive open-label TCZ. At this point, efficacy will be declared. Recruitment into the TCZ arm will continue until 300 patients have been enrolled.

The type I error rate will be controlled to ensure statistical validity is maintained. Specifically, the Lan-DeMets α -spending function that approximates the O'Brien-Fleming boundary will be applied to determine the critical value for stopping for positive efficacy at the interim analysis (DeMets and Lan 1994). Interim analyses for efficacy will use the Fisher's exact test for difference in proportions for mortality at 28 days and will utilize an O'Brien-Fleming alpha-spending function. The efficacy boundaries for the z-scores at the four looks (three interim analyses and primary analysis) are 4.364, 2.986, 2.377 and 2.011. The one-sided local significance levels at the three efficacy interim analyses are 0.000006392, 0.001415 and 0.008718.

Additional information regarding the efficacy interim analyses is detailed in the interim SAP. The critical value at the final analysis will be adjusted accordingly to maintain the protocol-specified overall type I error rate, per standard Lan-DeMets methodology.

The study management team will remain blinded unless the results meet the efficacy criteria (boundary is crossed). The Interim efficacy analyses will be produced by a statistical programmer and statistician independent of the study management team and will be reviewed by a DMC.

Full statistical details of the planned interim analyses, along with the rationale and timing will be documented in an interim statistical analysis plan, which will be made available to the relevant health authorities before the data snapshot for the first interim.

A DMC will also evaluate safety according to policies and procedures detailed in a DMC Charter. Regular safety reviews will begin after approximately 15 patients (10 TCZ,

5 placebo) have been enrolled and reached 14-day follow-up. Early stopping criteria based on compelling efficacy or an imbalance in adverse events will be detailed in the DMC charter. The safety interim analyses will also be conducted by a statistical programmer and statistician independent from the study management team and will be reviewed by the DMC. Interactions between the DMC and Sponsor will be carried out as specified in the DMC Charter.

The Data Monitoring Committee will initially consist of Sponsor representatives not directly involved in the SMT. If feasible during study conduct, the responsibilities may transition to a fully independent DMC (iDMC).

5. REFERENCES

- Bernard GR, Doig G, Hudson LD, et al. Quantification of organ failure for clinical trials and clinical practice. *Am J Respir Crit Care Med* 1995;151:Suppl:A323-A323
- DeMets DL, Lan KG. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13:1341-52.
- Sampson HA, Muñoz-Furlong A, Campbell RL, et al. Second symposium on the definition and management of anaphylaxis: summary report—Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *J Allergy Clin Immunol* 2006;117:391-7.
- Xu X, Han M, Li, T, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Submitted manuscript. [Resource on the internet]. 2020 [updated 5 March 2020; cited 17 March 2020]. <http://www.chinaxiv.org/abs/202003.00026>
- National Heart, Lung, and Blood Institute ARDS Clinical Trials Network, Truitt JD, Bernard GR, et al. Rosuvastatin for sepsis-associated acute respiratory distress syndrome. *N Engl J Med*. 2014;370(23):2191-2200.

Appendix 1 Protocol Synopsis

TITLE:	A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY TO EVALUATE THE SAFETY AND EFFICACY OF TOCILIZUMAB IN PATIENTS WITH SEVERE COVID-19 PNEUMONIA
PROTOCOL NUMBER:	WA42380
VERSION NUMBER:	3
EUDRACT NUMBER:	2020-001154-22
IND NUMBER:	148225
IND NUMBER:	NCT04320615
TEST PRODUCT:	Tocilizumab (RO4877533)
PHASE:	Phase III
INDICATION:	Severe COVID-19 pneumonia
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives and Endpoints

This study will evaluate the efficacy, safety, pharmacodynamics, and pharmacokinetics of tocilizumab (TCZ) compared with a matching placebo in combination with standard of care (SOC) in hospitalized patients with severe COVID-19 pneumonia. Specific objectives and corresponding endpoints for the study are outlined below.

Efficacy Objectives

Primary Efficacy Objective

The primary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

- Clinical status assessed using a 7-category ordinal scale at Day 28

Secondary Efficacy Objective

The secondary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia on the basis of the following endpoints:

- Time to clinical improvement (TTCI) defined as a National Early Warning Score 2 (NEWS2) of ≤ 2 maintained for 24 hours
- Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical status
- Incidence of mechanical ventilation
- Ventilator-free days to Day 28
- Incidence of intensive care unit (ICU) stay
- Duration of ICU stay

- Time to clinical failure, defined as the time to death, mechanical ventilation, ICU admission, or withdrawal (whichever occurs first). For patients entering the study already in ICU or on mechanical ventilation, clinical failure is defined as a one-category worsening on the ordinal scale, withdrawal or death.
- Mortality rate at Days 7, 14, 21, 28, and 60
- Time to hospital discharge or “ready for discharge” (as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or $\leq 2L$ supplemental oxygen)
- *Time to recovery, defined as discharged or “ready for discharge” (as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or $\leq 2L$ supplemental oxygen); OR, in a non-ICU hospital ward (or “ready for hospital ward”) not requiring supplemental oxygen*
- Duration of supplemental oxygen

Exploratory Efficacy Objective

The exploratory efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia on the basis of the following endpoints:

- Incidence of vasopressor use
- Duration of vasopressor use
- Incidence of extracorporeal membrane oxygenation (ECMO)
- Duration of ECMO
- Organ failure-free days to Day 28

Safety Objective

The safety objective for this study is to evaluate the safety of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia on the basis of the following endpoints:

- Incidence and severity of adverse events, with severity determined according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) v5.0
- SARS-CoV-2 (COVID-19) viral load over time, as collected by nasopharyngeal swab and bronchoalveolar lavage (BAL) samples (if applicable)
- Time to reverse-transcriptase polymerase chain reaction (RT-PCR) virus negativity
- The proportion of patients with any post-treatment infection
- Change from baseline in targeted clinical laboratory test results

Pharmacodynamic Objective

The pharmacodynamic objective for this study is to characterize the pharmacodynamic effects of TCZ in patients with COVID-19 pneumonia via longitudinal measures of the following analytes relative to baseline

- Serum concentrations of IL-6, sIL-6R, and CRP at specified timepoints

Pharmacokinetic Objective

The PK objective for this study is to characterize the TCZ PK profile in patients with COVID-19 pneumonia on the basis of the following endpoint:

- Serum concentration of TCZ at specified timepoints

Biomarker Objective

The exploratory biomarker objectives for this study are to identify and/or evaluate biomarkers that could be predictive of response to TCZ (i.e., predictive biomarkers), may serve as early surrogates of efficacy, may be associated with progression to a more severe disease state

(i.e., prognostic biomarkers), may be associated with susceptibility to developing adverse events or could lead to improved adverse event monitoring or investigation (i.e., safety biomarkers), could further evidence of TCZ pharmacological activity (i.e., pharmacodynamic biomarkers), and overall increase our knowledge and understanding of disease pathogenesis and drug safety, on the basis of the following endpoint:

- Assessments of individual biomarkers in relation to efficacy, safety, exposure and in both blood- and tissue-derived samples

STUDY DESIGN

Description of the Study

This is a Phase III, randomized, double-blind, placebo-controlled, multicenter study to assess the efficacy and safety of TCZ in combination with SOC compared with matching placebo in combination with SOC in hospitalized adult patients with severe COVID-19 pneumonia. The Sponsor intends to enroll approximately 450 patients that have been diagnosed with COVID-19 pneumonia and meet the entry criteria in centers globally.

Patients must be at least 18 years of age with confirmed SARS-CoV-2 (COVID-19) infection per WHO criteria, including a positive PCR of any specimen (e.g., respiratory, blood, urine, stool, other bodily fluid). At the time of enrollment, patients must have $\text{SpO}_2 \leq 93\%$ or $\text{PaO}_2/\text{FiO}_2 < 300$ mmHg despite being on SOC, which may include anti-viral treatment, low dose steroids, and supportive care.

Patients in whom, in the opinion of the treating physician, progression to death is imminent and inevitable within the next 24 hours, irrespective of the provision of treatments, will be excluded from the study. Patients with active tuberculosis (TB) or suspected active bacterial, fungal, viral, or other infection (besides COVID-19) will be excluded from the study.

Patients will be randomized as soon as possible after screening at a 2:1 ratio to receive blinded treatment of either TCZ or a matching placebo, respectively. Study treatment must be given in combination with SOC. The randomization will be stratified by geographic region (North America and Europe) and mechanical ventilation (yes, no). The proportion of patients who are on a mechanical ventilator at the time of randomization will be capped at no more than 50% of the overall study population.

Patients assigned to the TCZ arm will receive one infusion of TCZ 8 mg/kg, with a maximum dose of 800 mg, and patients assigned to the placebo arm will receive one infusion of placebo, both in addition to SOC.

For both arms, if the clinical signs or symptoms worsen or do not improve (reflected by sustained fever or at least a one-category worsening on the 7-category ordinal scale of clinical status), one additional infusion of blinded treatment of TCZ or placebo can be given, 8–24 hours after the initial infusion.

Patients who do not meet the criteria for participation in this study (screen failure) may qualify for one re-screening opportunity (for a total of 2 screenings per participant) at the investigator's discretion. Patients are not required to re-sign the consent form if they are re-screened within 7 days after previously signing the consent form.

The study assessments to be conducted include the following: physical examination, vital signs, oxygen saturation, assessment of consciousness, presence and absence of respiratory support, adverse events, concomitant therapies, clinical laboratory tests, and nasopharyngeal swabs.

Patients will be followed up for a total of 60 days after first dose of study medication.

If patients are discharged from hospital prior to Day 28, follow up visits should occur twice weekly through Day 28. Patients are encouraged to return for onsite visits at least weekly if possible; other follow-up visits may be conducted by telephone or home visits. Patients should return to the site for a Day 28 visit. After Day 28, all patients should have follow up visits on Day 35, Day 45, and Day 60; the Day 35 and Day 45 visits may be conducted by telephone or by home visits for discharged patients, while the Day 60 visit should be conducted onsite.

During the study, standard supportive care will be given according to clinical practice.

Number of Patients

This study aims to enroll approximately 450 hospitalized patients with severe COVID-19 pneumonia.

Target Population

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Signed Informed Consent Form by any patient capable of giving consent, or, when the patient is not capable of giving consent, by his or her legal/authorized representative
- Age ≥ 18 years at time of signing Informed Consent Form
- Ability to comply with the study protocol, in the investigator's judgment
- Hospitalized with COVID-19 pneumonia confirmed per a positive PCR of any specimen (e.g., respiratory, blood, urine, stool, other bodily fluid) and evidenced by chest X-ray or CT scan
- $SpO_2 \leq 93\%$ or $PaO_2/FiO_2 < 300$ mmHg

If a patient is on supplemental oxygen with $SpO_2 > 93\%$, but desaturation to $\leq 93\%$ on lower supplemental oxygen or ambient air is documented during screening, the inclusion criterion is met.

- For women of childbearing potential: agreement to remain abstinent (refrain from heterosexual intercourse) or use contraception, as defined below:

Women must remain abstinent or use contraceptive methods with a failure rate of $< 1\%$ per year during the treatment period and for 90 days after the final dose of study drug. Women must refrain from donating eggs during this same period.

A woman is considered to be of childbearing potential if she is postmenarchal, has not reached a postmenopausal state (≥ 12 continuous months of amenorrhea with no identified cause other than menopause), and is not permanently infertile due to surgery (i.e., removal of ovaries, fallopian tubes, and/or uterus) or another cause as determined by the investigator (e.g., Müllerian agenesis). The definition of childbearing potential may be adapted for alignment with local guidelines or regulations.

Examples of contraceptive methods with a failure rate of $< 1\%$ per year include bilateral tubal ligation, male sterilization, hormonal contraceptives that inhibit ovulation, hormone-releasing intrauterine devices, and copper intrauterine devices.

The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not adequate methods of contraception. If required per local guidelines or regulations, locally recognized adequate methods of contraception and information about the reliability of abstinence will be described in the local Informed Consent Form.

- For men: agreement to remain abstinent (refrain from heterosexual intercourse) or use a condom, and agreement to refrain from donating sperm, as defined below:

With a female partner of childbearing potential or pregnant female partner, men must remain abstinent or use a condom during the treatment period and for 60 days after the final dose of study drug to avoid exposing the embryo. Men must refrain from donating sperm during this same period.

The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not adequate methods of preventing drug exposure. If required per local guidelines or regulations, information about the reliability of abstinence will be described in the local Informed Consent Form.

Exclusion Criteria

Patients who meet any of the following criteria will be excluded from study entry:

- Known severe allergic reactions to TCZ or other monoclonal antibodies

- Active TB infection
- Suspected active bacterial, fungal, viral, or other infection (besides COVID-19)
- In the opinion of the investigator, progression to death is imminent and inevitable within the next 24 hours, irrespective of the provision of treatments
- Have received oral anti-rejection or immunomodulatory drugs (including TCZ) with the past 3 months
- Participating in other drug clinical trials (participation in COVID-19 anti-viral trials may be permitted if approved by Medical Monitor)
- ALT or AST > 10 x ULN detected within 24 hours at screening (according to local laboratory reference ranges)
- ANC < 1000/ μ L at screening (according to local laboratory reference ranges)
- Platelet count < 50,000/ μ L at screening (according to local laboratory reference ranges)
- Pregnant or breastfeeding, or positive pregnancy test in a pre-dose examination
- Treatment with an investigational drug within 5 half-lives or 30 days (whichever is longer) of randomization (investigational COVID-19 antivirals may be permitted after consultation with the Medical Monitor)
- Any serious medical condition or abnormality of clinical laboratory tests that, in the investigator's judgment, precludes the patient's safe participation in and completion of the study

End of Study

The end of this study is defined as the date when the last patient, last visit (LPLV) occurs or the date at which the last data point required for last follow-up is received from the last patient, whichever occurs later. The end of the study is expected to occur approximately 2 months after the last patient is enrolled.

In addition, the Sponsor may decide to terminate the study at any time.

Length of Study

The total length of the study, from screening of the first patient to the end of the study, is expected to be approximately 6 months.

Investigational Medicinal Products

Test Product (Investigational Drug)

Patients assigned to the active arm will receive one or two doses of tocilizumab (TCZ) via IV infusion at a dose of 8 mg/kg IV to a maximum of 800 mg per dose.

Comparator

Patients assigned to the comparator arm will receive one or two doses of placebo via IV.

Statistical Methods

Primary Analysis

The primary efficacy objective for this study is to evaluate the efficacy of TCZ plus standard of care, compared with placebo plus standard of care using the following endpoint:

- Clinical status assessed using a 7-category ordinal scale at Day 28

Assessment of patient status using an ordinal scale will be recorded at baseline and daily in the morning (between 8 am and 12 pm) while hospitalized. The ordinal scale categories are as follows:

1. Discharged (or "ready for discharge" as evidenced by normal body temperature and respiratory rate, and stable oxygen saturation on ambient air or \leq 2L supplemental oxygen)
2. Non-ICU hospital ward (or "ready for hospital ward") not requiring supplemental oxygen
3. Non-ICU hospital ward (or "ready for hospital ward") requiring supplemental oxygen
4. ICU or non-ICU hospital ward, requiring non-invasive ventilation or high-flow oxygen
5. ICU, requiring intubation and mechanical ventilation

6. ICU, requiring ECMO or mechanical ventilation and additional organ support (e.g., vasopressors, renal replacement therapy)

7. Death

The estimand is the difference in distributions between TCZ plus SOC and placebo plus SOC, which will be tested using a non-parametric method, the Van Elteren test, including the stratification factors at randomization (region [North America, Europe] and mechanical ventilation [yes, no]). The median ordinal scale result for each treatment group and the corresponding 95% CI for the median will be presented along with the Van Elteren p-value, as well as the difference in medians and a 95% CI for the difference.

Further details of the primary endpoint analysis will be included in the SAP.

As an additional analysis, the clinical status according to the 7-category ordinal scale will be compared between the TCZ group and the placebo group at Day 28, using a proportional odds model accounting for stratification factors at randomization in the model (region [North America, Europe] and mechanical ventilation [yes, no]). The odds ratio, p-value, and 95% confidence interval will be presented.

For patients who withdraw before Day 28, their last post baseline ordinal category prior to withdrawal will be used in the analysis. Any other missing data handling rules for the primary endpoint will be specified in the SAP.

Determination of Sample Size

The estimated sample size was determined for the primary endpoint of comparison of clinical status based on a 7-category ordinal scale at Day 28 using *the Van Elteren test*.

The total mITT sample size of 450 with a 2:1 randomization of TCZ to placebo patients provides approximately 90% power to detect a *difference in distribution between the treatment groups of the ordinal scale at Day 28 using a two-sided Van Elteren test at the 5% significance level* under the following assumptions of the expected probability distribution of patients in the placebo arm:

1 (discharge)	2	3	4	5	6	7 (death)
0.58	0.05	0.09	0.09	0.02	0.02	0.15

And, assuming proportional odds with an odds ratio of 2, the expected distribution in the TCZ arm would be:

1 (discharge)	2	3	4	5	6	7 (death)
0.734	0.039	0.064	0.058	0.012	0.012	0.081

In addition, this sample size provides approximately 90% power to detect a ratio of 2 (TCZ to PBO) for the odds of being in a category or a better using a proportional odds model with a two-sided p-value at the 5% significance level.

Assuming proportional odds and the given distribution of the placebo group, the smallest odds ratio that could be statistically significant would be approximately 1.5.

This sample size also provides *approximately 90% power to detect a 10% absolute difference in mortality rate under the assumption of a 15% mortality rate in the placebo group.*

Planned Interim Analyses

Up to three interim looks for efficacy will be carried out on the data with mortality rate at 28 days (secondary endpoint) evaluated for interim efficacy analyses. The interim looks will occur after roughly 111, 222, and 333 patients are enrolled, but all interims are subject to change depending on enrollment. If the sample size is increased during the study, the remaining efficacy interims will be performed at similar proportions of information to the original planned efficacy interim analyses.

The first efficacy interim analysis will be conducted when approximately 111 patients (74 TCZ and 37 placebo) have reached the 28-day follow-up time point and will be based on the mortality rate at 28 days (secondary endpoint). If the results of one of the interim analyses meets the pre-specified criteria for efficacy, further enrollment in the placebo arm will be discontinued and all enrolled patients will receive open-label TCZ. At this point, efficacy will be

declared. *If the study is at least 90% enrolled within 5 weeks (28 days follow up plus 1 week to perform the analysis) of the 111th patient being enrolled, then no interim analyses will be conducted.*

If there is a potential for further recruitment into the placebo arm to be stopped for positive efficacy because of the interim analysis, the type I error rate will be controlled to ensure statistical validity is maintained. Specifically, the Lan-DeMets α -spending function that approximates the O'Brien-Fleming boundary will be applied to determine the critical value for stopping for positive efficacy at the interim analysis (DeMets and Lan 1994). Interim analyses for efficacy will use the Fisher's exact test for difference in proportions and will utilize an O'Brien-Fleming alpha-spending function. The efficacy boundaries for the z-scores at the four looks (three interim looks and final analysis) are 4.364, 2.986, 2.377, and 2.011.

The critical value at the final analysis will be adjusted accordingly to maintain the protocol-specified overall type I error rate, per standard Lan-DeMets methodology.

The study management team will remain blinded unless the results meet the efficacy criteria. The interim efficacy analyses will be produced by a statistical programmer independent of the study management team and will be reviewed by a Data Monitoring Committee (DMC).

Full statistical details of the planned interim analyses, along with the rationale and timing will be documented in an interim statistical analysis plan, which will be made available to the relevant health authorities before the data snapshot for the first interim.

A Data Monitoring Committee will also evaluate safety according to policies and procedures detailed in a DMC Charter. Regular safety reviews will begin after approximately 15 patients (10 TCZ, 5 placebo) have been enrolled and reached 14-day follow-up. Early stopping criteria based on compelling efficacy or an imbalance in adverse events will be detailed in the DMC charter. The safety interim analyses will also be conducted by a statistical programmer independent from the study management team and will be reviewed by the DMC. Interactions between the DMC and Sponsor will be carried out as specified in the DMC Charter.

The Data Monitoring Committee may initially consist of Sponsor representatives not involved in any operational aspects of the study before transitioning to a fully independent data monitoring committee (iDMC) when feasible.

Appendix 2 Schedule of Activities: Days 1 and 2

	Screening ^{a, b}	Baseline			
Study Day	-2 to 0	1		2	
<i>Time Post Initial Treatment (Assessment Window)</i>		0 <i>Pre-dose (-4 hrs)</i>	<i>15 min After end of infusion (+1 hr)</i>	<i>24 hrs (±4 hrs)</i>	<i>36 hrs (±4 hrs)</i>
Informed consent	x				
Inclusion/exclusion criteria	x	x			
Demographic data	x				
Randomization		x			
Medical history		x			
Complete physical examination ^c	x				
Weight		x			
COVID-19 diagnosis ^d	x				
Chest X-ray/CT scan ^e	x				
ECG	x				
Pregnancy test ^f	x				
PaO ₂ /FiO ₂ ^g	x	<i>← Optional →</i>			
SpO ₂ ^h	x	x	x	x	x
Vital signs ^h	x	x	x	x	x
Ordinal scoring ⁱ		x		x	
Adverse events ^j		x		x	
Concomitant medications ^k		x		x	
Hematology ^l	x	x		x	
Chemistry ^m	x	x		x	

Appendix 2 Schedule of Activities: Days 1 and 2 (cont.)

	Screening ^{a, b}	Baseline			
Study Day	-2 to 0	1		2	
<i>Time Post Initial Treatment (Assessment Window)</i>		0 <i>Pre-dose (-4 hrs)</i>	<i>15 min After end of infusion (+1 hr)</i>	<i>24 hrs (±4 hrs)</i>	<i>36 hrs (±4 hrs)</i>
Study drug administration ⁿ		x			
Central Labs					
Serum PD (CRP, IL-6, sIL-6R)		x ^o	x ^o	x	x
Serum PK ^p		x ^q	x ^q	x	x
Serum sample for exploratory biomarkers		x		x	
SARS-CoV-2 viral load ^r		x		x	
Serum SARS-CoV-2 antibody titer		x			
Cryopreserved PBMCs ^s		x		x	
Whole blood in PAXgene [®] tubes for RNA analyses ^t		x			

CRP = c-reactive protein; CT = computed tomography; ECG = electrocardiogram; eCRF = electronic case report form; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events v.5.0; NEWS2 = National Early Warning Score; PaO₂/FiO₂ = arterial oxygen partial pressure/fraction of inspired oxygen; PBMCs = peripheral blood mononuclear cells; PK = pharmacokinetic; PRO-CTCAE = NCI Patient-Reported Outcomes Common Terminology Criteria for Adverse Events; SpO₂ = peripheral capillary oxygen saturation.

Note: On treatment days, all assessments should be performed prior to dosing, unless otherwise specified.

- ^a Results from standard-of-care tests or examinations performed prior to obtaining informed consent and within 48 hours before randomization may be used; such tests do not need to be repeated for screening.
- ^b Informed consent must be documented before any study-specific screening procedure is performed.
- ^c A complete physical examination, performed at screening and per the investigator's discretion during the study, includes an evaluation of the head, eyes, ears, nose, and throat, and the cardiovascular, dermatologic, musculoskeletal, respiratory, gastrointestinal, genitourinary, and neurologic systems. Any abnormality identified at screening should be recorded on the General Medical History and Baseline Conditions eCRF. New or worsened clinically significant abnormalities identified during the study should be reported as adverse events.
- ^d COVID-19 test (SARS-CoV2 PCR) to confirm diagnosis should be performed within 7 days of randomization.

Appendix 2 Schedule of Activities: Days 1 and 2 (cont.)

- e Screening chest X-ray or CT scans should be performed within 48 hours prior to randomization. If additional chest X-rays/CT scans are taken per local practice during the study, this information should be provided in the eCRF.
- f For women of childbearing potential, including those who have had a tubal ligation, positive urine test results will be confirmed with a serum pregnancy test. Study drug infusion must not be administered unless the serum pregnancy test result is negative.
- g If arterial blood gases are measured.
- h All vital sign measurements (i.e., respiratory rate, pulse rate, systolic and diastolic blood pressures, and body temperature), oxygen saturation and NEWS2-specific assessments (i.e., consciousness and presence or absence of oxygen support) should be recorded together twice daily with approximately 12 hours in between while the patient remains hospitalized. If measured more than once during a 12-hour period, the worst values (highest temperature, respiratory rate, and heart rate; lowest blood pressure, oxygen saturation, and consciousness level) during that period should be recorded on the eCRF.
- i Assessment of clinical status using the ordinal scale should be recorded at baseline on Day 1/Visit 1 then again daily every morning (between 8 am and 12 pm) for patients who remain hospitalized.
- j After informed consent has been obtained but prior to initiation of study drug, only serious adverse events caused by a protocol-mandated intervention should be reported. After initiation of study drug, all adverse events will be reported until 60 days after the final dose of study drug. After this period, the Sponsor should be notified if the investigator becomes aware of any serious adverse event that is believed to be related to prior study drug treatment.
- k Medication (e.g., prescription drugs, over-the-counter drugs, vaccines, herbal or homeopathic remedies, nutritional supplements) used by a patient in addition to protocol-mandated treatment from 7 days prior to initiation of study drug to the study completion/discontinuation visit.
- l Hematology includes WBC count, RBC count, hemoglobin, hematocrit, platelet count, differential count (neutrophils, eosinophils, basophils, monocytes, lymphocytes), and lymphocyte subsets (T cells, B cells, and NK cells; if the test is available at the site).
- m Chemistry panel (serum or plasma) includes bicarbonate or total carbon dioxide (if considered standard of care for the region), sodium, potassium, chloride, glucose, BUN or urea, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, ALP, ALT, and/or AST, uric acid, LDH, ferritin, and D-dimer.
- n Study drug should be administered after collection of all samples for pharmacodynamic and exploratory biomarker analyses. The initial study drug infusion should be given within 4 hours of randomization. If the clinical signs or symptoms worsen or do not improve (reflected by sustained fever or at least a one-category worsening on the 7-category ordinal scale of clinical status), one additional infusion of blinded treatment of TCZ or placebo can be given, 8–24 hours after the initial infusion.
- o On Day 1, CRP, IL-6, and sIL-6R samples should be drawn 0–4 hours before the start of infusion and then within 15 minutes (to 1 hour) after the end of the infusion, on the opposite arm as the infusion. Patients receiving a second infusion of study drug should provide extra samples for CRP, IL-6, and sIL-6R prior to and 15 minutes after the end of the infusion, on the opposite arm as the infusion.
- p Patients receiving a second infusion of study drug should provide an extra PK sample prior to and 15 minutes (to 1 hour) after the end of the infusion, on the opposite arm as the infusion.

Appendix 2 Schedule of Activities: Days 1 and 2 (cont.)

- ¶ On Day 1, PK samples should be drawn 0–4 hours before the start of infusion and then within 15 minutes (to 1 hour) after the end of the infusion, on the opposite arm as the infusion.
- ¶ Viral load will be assessed by nasopharyngeal swab. Patients who are intubated and undergo bronchoalveolar lavage will have samples taken for virological assessment. It may only be possible to access one nostril if a patient has a nasogastric tube in place, in which case sites may collect a sample from one nostril only and where possible the same nostril should be used.
- ¶ For sites capable of sample collection for analysis of T cells by high-dimensional cytometry.
- ¶ The first draw of blood should not be for PAXgene® tubes to avoid contact with RNA preservation reagent inside the tube.

Appendix 3 Schedule of Activities: Days 3–28

	Days 3–28 ^a																										Study Completion/ Discontinuation	
Study Day	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28		
Chest X-ray/CT scan					x							x							x							x	x	
Vital signs ^b	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
PaO ₂ /FiO ₂ ^c	← Optional →																										Optional	
SpO ₂ ^b	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Ordinal scoring ^d	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Adverse events ^e	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Concomitant medications ^f	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Hematology ^g	x				x			x				x							x								x	x
Chemistry ^h	x				x			x				x							x								x	x
<i>Central Labs</i>																												
Serum PD (CRP, IL-6, sIL-6R)	x				x							x							x								x	x
Serum PK					x							x							x								x	x
Serum sample for exploratory biomarkers	x				x							x							x								x	x
SARS-CoV-2 viral load ⁱ	x	x	x	x	x			x				x							x								x	x
Serum SARS-CoV-2 antibody titer																											x	x

Appendix 3 Schedule of Activities: Days 3–28 (cont.)

	Days 3–28 ^a																								Study Completion/ Discontinuation		
Study Day	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
Cryopreserved PBMCs ^j	x				x							x							x							x	x
Whole blood in PAXgene [®] tubes for RNA analyses ^k	x				x																					x	x

BAL = bronchoalveolar lavage; CRP = c-reactive protein; CT = computed tomography; ECG = electrocardiogram; eCRF = electronic case report form; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events v.5.0; PaO₂/FiO₂ = arterial oxygen partial pressure/fraction of inspired oxygen; PBMCs = peripheral blood mononuclear cells; PK = pharmacokinetic; PRO-CTCAE = NCI Patient-Reported Outcomes Common Terminology Criteria for Adverse Events; SpO₂ = peripheral capillary oxygen saturation..

Note: For patients who have been discharged, all assessments should be performed within ±3 days of the scheduled onsite visit.

- ^a If patients are discharged from hospital prior to Day 28, follow-up visits should occur twice weekly through Day 28. Patients are encouraged to return for onsite visits at least weekly if possible; other follow-up visits may be conducted by telephone or home visits. Patients should return to the site for a Day 28 visit.
- ^b All vital sign measurements (i.e., respiratory rate, pulse rate, systolic and diastolic blood pressures, and body temperature), oxygen saturation and NEWS2-specific assessments (i.e., consciousness and presence or absence of oxygen support) should be recorded together twice daily with approximately 12 hours in between while the patient remains hospitalized. If measured more than once during a 12-hour period, the worst values (highest temperature, respiratory rate, and heart rate; lowest blood pressure, oxygen saturation, and consciousness level) during that period should be recorded on the eCRF. Following hospital discharge these parameters should be recorded once at each return visit to the clinic. Vital signs and oxygen saturation will not be recorded if follow-up visits are conducted by telephone.
- ^c If arterial blood gases are measured.
- ^d Assessment of clinical status using the ordinal scale should be recorded at baseline on Day 1/Visit 1 then again daily every morning (between 8 am and 12 pm) for patients who remain hospitalized.
- ^e After informed consent has been obtained but prior to initiation of study drug, only serious adverse events caused by a protocol-mandated intervention should be reported. After initiation of study drug, all adverse events will be reported until 60 days after the final dose of study drug. After this period, the Sponsor should be notified if the investigator becomes aware of any serious adverse event that is believed to be related to prior study drug treatment.

Appendix 3 Schedule of Activities: Days 3–28 (cont.)

- ^f Medication (e.g., prescription drugs, over-the-counter drugs, vaccines, herbal or homeopathic remedies, nutritional supplements) used by a patient in addition to protocol-mandated treatment from 7 days prior to initiation of study drug to the study completion/discontinuation visit.
- ^g Hematology includes WBC count, RBC count, hemoglobin, hematocrit, platelet count, differential count (neutrophils, eosinophils, basophils, monocytes, lymphocytes), and lymphocyte subsets (T cells, B cells, and NK cells; if the test is available at the site).
- ^h Chemistry panel (serum or plasma) includes bicarbonate or total carbon dioxide (if considered standard of care for the region), sodium, potassium, chloride, glucose, BUN or urea, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, ALP, ALT, and/or AST, uric acid, LDH, ferritin, and D-dimer.
- ⁱ Viral load will be assessed by nasopharyngeal swab. Patients who are intubated and undergo bronchoalveolar lavage will have samples taken for virological assessment. It may only be possible to access one nostril if a patient has a nasogastric tube in place, in which case sites may collect a sample from one nostril only and where possible the same nostril should be used.
- ^j For sites capable of sample collection for analysis of T cells by high-dimensional cytometry.
- ^k The first draw of blood should not be for PAXgene[®] tubes to avoid contact with RNA preservation reagent inside the tube.

Appendix 4 Schedule of Activities: After Day 28

Study Day (Assessment Window)	Study Completion		
	35 ^a (±3 days)	45 ^a (±3 days)	60 (±3 days)
Chest X-ray/CT scan			x
SARS-CoV-2 viral load ^b	x	x	x
Vital signs ^c	x	x	x
SpO ₂ ^c	x	x	x
Ordinal scoring ^d	x	x	x
Adverse events ^e	x	x	x
Concomitant medications ^f	x	x	x
Hematology ^g	x	x	x
Chemistry ^h	x	x	x
Central Labs			
Serum PD (CRP, IL-6, sIL-6R)	x		x
Serum PK	x		x
Serum sample for exploratory biomarkers	x		x
Serum SARS-Cov-2 antibody titer			x

CRP=c-reactive protein; CT=computed tomography; PK=pharmacokinetic; SpO₂=peripheral capillary oxygen saturation.

- ^a If patients are unable to return for onsite visits at Day 35 and/or Day 45, these may be conducted by telephone or home visits. Patients should return to the site for a Day 60 Study Completion visit.
- ^b Patients who remain in hospital will have viral load assessed by nasopharyngeal swabs; these will be done if there is evidence of on-going infection.
- ^c For patients who remain in hospital, vital sign measurements and NEWS2-specific assessments should be conducted twice daily. Following hospital discharge, these parameters should be recorded once at each return visit to the clinic. Vital signs and oxygen saturation will not be recorded if follow-up visits are conducted by telephone.

Appendix 4 Schedule of Activities: After Day 28 (cont.)

- ^d Assessment of clinical status using the ordinal scale should be recorded at baseline on Day 1/Visit 1 then again daily every morning (between 8 am and 12 pm) for patients who remain hospitalized.
- ^e After informed consent has been obtained but prior to initiation of study drug, only serious adverse events caused by a protocol-mandated intervention should be reported. After initiation of study drug, all adverse events will be reported until 60 days after the final dose of study drug. After this period, the Sponsor should be notified if the investigator becomes aware of any serious adverse event that is believed to be related to prior study drug treatment.
- ^f Medication (e.g., prescription drugs, over-the-counter drugs, vaccines, herbal or homeopathic remedies, nutritional supplements) used by a patient in addition to protocol-mandated treatment from 7 days prior to initiation of study drug to the study completion/discontinuation visit.
- ^g Hematology includes WBC count, RBC count, hemoglobin, hematocrit, platelet count, differential count (neutrophils, eosinophils, basophils, monocytes, lymphocytes), and lymphocyte subsets (T cells, B cells, and NK cells). Hematology labs will not be performed if follow-up visits are conducted by telephone.
- ^h Chemistry panel (serum or plasma) includes bicarbonate or total carbon dioxide (if considered standard of care for the region), sodium, potassium, chloride, glucose, BUN or urea, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, ALP, ALT, and/or AST, uric acid, LDH, ferritin, and D-dimer. Chemistry labs will not be performed if follow-up visits are conducted by telephone.

Appendix 5 National Early Warning Score 2 (NEWS2)

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	

SpO₂ = oxygen saturation.

The oxygen saturation should be scored according to either the SpO₂ Scale 1 or 2 presented in the table above. The SpO₂ Scale 2 is for patients with a target oxygen saturation requirement of 88%–92% (e.g., in patients with hypercapnic respiratory failure related to advanced lung diseases, such as chronic obstructive pulmonary disease [COPD]). This should only be used in patients confirmed to have hypercapnic respiratory failure by blood gas analysis on either a prior or their current hospital admission.

The decision to use the SpO₂ Scale 2 should be made by the treating physician and should be recorded in the eCRF. In all other circumstances, the SpO₂ Scale 1 should be used.

For physiological parameter “Air or Oxygen?”: Any patients requiring the use of oxygen or other forms of ventilation to maintain oxygen saturations and support respiration should be assigned a score of 2.

The consciousness level should be recorded according to the best clinical condition of the patient during the assessment. Patients who are assessed as “Alert” (A) should be assigned a score of 0. Patients assessed as “New Confusion” (C), “Responsive to Voice” (V), “Responsive to Pain” (P), or “Unconscious” should be assigned a score of 3.

Appendix 5 National Early Warning Score 2 (NEWS2) (cont.)

Scores should be assigned for respiratory rate, systolic blood pressure, pulse, and temperature according to the table above.

NEWS2 values will be calculated electronically throughout the study by the Sponsor based upon entry of vital sign parameters by the investigator in the appropriate eCRF.

Example Case Calculation:

An 82-year-old lady was admitted, tested positive to COVID-19 and admitted to high dependency unit for non-invasive ventilation. Her taken observations and corresponding NEWS2 score are as follows:

Physiological Parameter	Observation	Component Score
Respiratory rate (per min)	26	3
Oxygen saturation (SpO ₂ %)	95%	1
Supplemental Oxygen	Yes	2
Systolic blood pressure (mmHg)	95	2
Pulse Rate (bpm)	109	1
Conscious level	New confusion	3
Temperature (°C)	39	1
	Total NEWS2 Score	13

REFERENCE

Royal College of Physicians. National early warning score (NEWS) 2. Standardizing the assessment of acute-illness severity in the NHS. London: RCP, 2017.

Appendix 6 Cochran-Mantel-Haenszel Test

- The weighted difference in proportions is the difference in the response rates in the experimental treatment group compared with the control treatment group, adjusted for any stratification factors. With two stratification factors, the number of patients in each strata is defined as n_{ijk} where i is the level of the first stratification factor and j is the level of the second stratification factor and k is treatment group (experimental or control). The number of events in each strata is denoted by x_{ijk} , where i , j and k are as above. The proportion of responders in each strata will be calculated by:

$$p_{ijk} = \frac{x_{ijk}}{n_{ijk}} \text{ where } i, j \text{ and } k \text{ are as above}$$

- The difference in proportions for each strata will then be calculated as the proportion of patients in each strata in the experimental treatment group (EXP) minus the proportion of patients in each strata in the control treatment group (CON) and denoted

$$d_{ij} = p_{ijEXP} - p_{ijCON}, \text{ for } i \text{ and } j \text{ as above.}$$

- The weights for each strata (i , j) will be calculated as follows:

$$w_{ij} = \frac{n_{ijEXP} * n_{ijCON}}{n_{ijEXP} + n_{ijCON}}$$

- Within each strata, the weighted differences in the proportions in each of the treatment groups will be calculated as follows:

$$wd_{ij} = w_{ij}d_{ij}$$

- and then summed:

$$WD = \sum_i \sum_j wd_{ij}$$

- After calculation of the weighted difference in proportions, the calculation of the 95% confidence interval is as follows;
- Continuity-corrected Proportions

$$p_{ijk}^{\#} = \frac{x_{ijk} + 0.5}{n_{ijk} + 1}$$

- Variances

$$Upvar_{ij} = w_{ij}^2 \left[p_{ijEXP}^{\#} \frac{(1 - p_{ijEXP}^{\#})}{n_{ijEXP}} + p_{ijCON}^{\#} \frac{(1 - p_{ijCON}^{\#})}{n_{ijCON}} \right]$$

Appendix 6 Cochran-Mantel-Haenszel Test (cont.)

- To calculate the sum of the weights and variances over all strata:

Sum over Strata

$$W = \sum_i \sum_j w_{ij} \quad (\text{sum of weights})$$

$$Var = \sum_i \sum_j Up var_{ij} \quad (\text{sum of variances})$$

Point Estimate and Standard Error

$$d = \frac{WD}{W} ; \quad se = \sqrt{\frac{Var}{W^2}}$$

Stratified 95% Confidence Intervals

$$\text{Lower Limit} = d - 1.96se$$

$$\text{Upper Limit} = d + 1.96se$$