STATISTICAL ANALYSIS PLAN

TITLE:	**A PHASE III, RANDOMIZED, DOUBLE-BLIND, MULTICENTER STUDY TO EVALUATE THE EFFICACY AND SAFETY OF REMDESIVIR PLUS TOCILIZUMAB COMPARED WITH REMDESIVIR PLUS PLACEBO IN HOSPITALIZED PATIENTS WITH SEVERE COVID-19 PNEUMONIA**
PROTOCOL NUMBER:	WA42511
STUDY DRUG:	Tocilizumab (RO4877533), Remdesivir
VERSION NUMBER:	2
IND NUMBER:	148225
EUDRACT NUMBER:	2020-002275-34
SPONSOR:	F. Hoffmann-La Roche Ltd.
PLAN PREPARED BY:	[redacted], Ph.D (Version 1), [redacted], M.S. (Version 1), [redacted], Ph.D (Version 2)
DATE FINAL:	Version 1: 4 December 2020
DATE AMENDED:	See electronic date stamp below

Date and Time(UTC)	Reason for Signing	Name
22-Feb-2021 20:41:04	Company Signatory	[redacted]

STATISTICAL ANALYSIS PLAN APPROVAL

CONFIDENTIAL

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

This Statistical Analysis Plan Version 2 was amended from Version 1 to address health authority feedback as follows:

- The derivation of the primary endpoint was updated
- Two additional secondary endpoints were added:
- Proportion of patients discharged or “ready for discharge” up to Day 28
- Proportion of patients who require initiation of mechanical ventilation postbaseline or die up to Day 28 and up to Day 60
- The definition of the analysis populations was clarified
- Study Day 1 was changed to the day of randomization rather than the day of first TCZ/PBO administration
- Time to event endpoints were changed from “time from administration of TCZ/PBO” to “time from randomization”
- The censoring rules and analysis approach for the time to event endpoints were updated
- The analysis method for the duration and ordinal endpoints was updated
- Tipping point analyses were added for the primary and key secondary endpoints
- The PD analyses section was removed

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 9

2.1 Protocol Synopsis 11

2.2 Endpoints 11

2.2.1 Primary Efficacy Endpoints 11

2.2.2 Secondary Efficacy Endpoints 11

2.2.3 Exploratory Efficacy Endpoints 12

2.2.4 Biomarkers 13

2.2.5 Safety Endpoints 13

2.3 Determination of Sample Size 13

2.4 Analysis Timing 13

3. STUDY CONDUCT 14

3.1 Randomization, Stratification and Blinding 14

3.2 Data Monitoring 15

4. STATISTICAL METHODS 16

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 17

4.3.1 Demographics 17

4.3.2 Baseline Characteristics 17

4.3.3 Medical History 18

4.3.4 Previous and Concomitant Medications 18

4.4 Visit Labels 18

4.5 Efficacy Analysis 21

4.5.1 Primary Efficacy Endpoint 21

4.5.2 Controlling for Type I Error 23

4.5.3 Type I Error Controlled Secondary Endpoints 24

4.5.3.1 Time to Mechanical Ventilation or Death up to Day 28 24

4.5.3.2 Clinical Status on Day 14 25
4.5.3.3 Time to Death up to Day 28 27
4.5.4 Other Secondary Endpoints 28
4.5.4.1 Time to Event Analyses 28
4.5.4.2 Incidence Endpoints 30
4.5.4.3 Ordinal endpoints 32
4.5.4.4 Duration endpoints 32
4.5.4.5 Meta-Analysis 33
4.5.5 Exploratory Efficacy Endpoints 33
4.5.6 Subgroup Analyses 37
4.5.7 Sensitivity and Tipping Point Analyses 37
4.5.7.1 Prognostic Covariates 37
4.5.7.2 Clinical Status at Day 14 37
4.5.7.3 Primary Time to Event Endpoint 38
4.5.7.4 Time to Death and Time to Mechanical Ventilation or Death up to Day 28 41
4.6 Safety Analyses 41
4.6.1 Exposure to Study Medication 41
4.6.2 Adverse Events 42
4.6.3 Laboratory Data 43
4.6.4 Vital Signs 43
4.7 Interim Analyses 44
5. REFERENCES 44

LIST OF TABLES

Table 1 Time Windows for Assigning Assessment Study Days to Study Visits Labels for Deaths and Discharge 20
Table 2 Time Windows for Assigning Assessment Study Days to Study Visits for Hematology Parameters 20
Table 3 Time to Hospital Discharge or “Ready for Discharge” Censoring Rules 23
Table 4 Time to Mechanical Ventilation or Death Censoring Rules 25
Table 5 Time to Death Censoring Rules 28
Table 6 Time to Improvement in Ordinal Clinical Status Censoring Rules 30

LIST OF FIGURES

Figure 1 Study WA42511 Schema 11

LIST OF APPENDICES

Appendix 1 Protocol Synopsis 46
Appendix 2 Schedule of Activities: Screening, Days 1 and 2 53
Appendix 3 Schedule of Activities: Days 3–28 57
Appendix 4 Schedule of Activities: After Day 28 60
Appendix 5 National Early Warning Score 2 (NEWS2) 63

GLOSSARY OF ABBREVIATIONS

AE	adverse event
AEGT	adverse event grouped term
AESI	adverse event of special interest
BAP	biomarker analysis plan
CI	confidence interval
CIF	cumulative incidence function
CMH	Cochran-Mantel-Haenszel
COVID-19	Corona Virus Disease-19
CRP	C-reactive protein
CT	computed tomography
DMC	data monitoring committee
ECG	electrocardiogram
ECMO	extracorporeal membrane oxygenation
eCRF	electronic Case Report Form
ICU	intensive care unit
iDMC	independent Data Monitoring Committee
IL-6	interleukin 6
iSAP	interim Statistical Analysis Plan
IV	intravenous
IxRS	interactive voice or web-based response system
MAP	meta-analysis Statistical Analysis Plan
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 partial pressure of oxygen (PaO_2) and fraction of inspired oxygen, (FiO_2)
PBO	placebo
PD	pharmacodynamics
PI	probabilistic index
PK	pharmacokinetic
PT	preferred term
RDV	remdesivir
SAE	serious adverse event
SAP	Statistical Analysis Plan
SMQ	Standard MedDRA Query

SMT	study management team
SOC	standard of care
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 WA42511. Any analyses of biomarkers or virologic data will be covered by a separate analysis plan. Analyses of pharmacokinetic data will be covered by a separate analysis plan.
A meta-analysis of this study along with data from Studies WA42380 and ML42528 for mortality may be conducted, along with the primary and selected secondary endpoints and this analysis will be specified in a separate meta-analysis Statistical Analysis Plan (MAP, see Section 4.5.4.5).

At the start of the pandemic in China, physicians in China initiated the off-label use of tocilizumab (TCZ) in the treatment of patients with COVID-19 pneumonia. The first study reported was a 21-patient retrospective study in which patients with severe or critical COVID-19 pneumonia were treated with TCZ (Xu et al. 2020).

Patients received TCZ 400 mg intravenous (IV) plus standard of care (SOC). The average age of the patients was 56.8 (±16.5) years, ranging from 25 to 88 years. Seventeen patients (81.0%) were assessed as severe and 4 patients (19.0%) were assessed as critical. Most patients (85.0%) presented with lymphopenia. The C-reactive protein (CRP) levels were increased in all 20 patients evaluated (mean: 75.06 [±66.80] mg/L). The median procalcitonin value was 0.33 (±0.78) ng/mL, and only 2 of 20 patients (10.0%) presented with an abnormal value. Mean modified intent to treat (interleukin-6 [IL-6]) level before TCZ was 132.38 (±278.54) pg/mL (normal < 7 pg/mL).

SOC consisted of lopinavir, methylprednisolone, other symptom relievers, and oxygen therapy as recommended by the Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial Version 6) (China National Health Commission 2020). All 21 patients had received routine SOC treatment for a week before deteriorating with sustained fever, hypoxemia, and chest computed tomography (CT) image worsening.

According to the authors, after TCZ treatment, fever returned to normal and all other symptoms improved remarkably. Fifteen of the 20 patients (75.0%) had lowered their oxygen intake, and 1 patient needed no oxygen therapy. The CT scans showed significant remission of opacities in both lungs in 19 of 20 patients (90.5%) after treatment with TCZ. No adverse drug reactions and no subsequent pulmonary infections were reported.

Nineteen patients (90.5%) were discharged at the time of the report, including 2 critical patients. There were no deaths among the 21 treated patients.

The study authors concluded that TCZ is an effective treatment for patients with severe COVID-19 pneumonia (Xu et al. 2020).

On 3 March 2020, TCZ was included in the seventh updated diagnosis and treatment plan for COVID-19 issued by the China National Health Commission as one treatment option for patients with severe or critical forms of COVID-19 pneumonia.

Since that initial report, multiple investigator-initiated studies of TCZ for the treatment of COVID-19 pneumonia have been published (Gupta et al. 2020; Guaraldi et al. 2020; Hermine et al. 2020; Perrone et al. 2020; Salvarani et al. 2020; Stone et al. 2020; Guaraldi et al. 2020). Results have been mixed, with some evidence of benefit in patients with severe disease at baseline but no benefit in moderate COVID-19.

In addition, two Sponsor-initiated studies have been conducted. A randomized, double-blind, Phase III study (WA42380) evaluating the safety and efficacy of TCZ administered intravenously compared with placebo plus SOC therapy in hospitalized patients with severe COVID-19 pneumonia read out in August 2020 (NCT04320615). Though the primary endpoint did not meet statistical significance, a decrease in time to discharge, and a reduction in duration of intensive care unit (ICU) stay were observed. A second trial in milder patients with COVID-19 (Study ML42528) met the primary endpoint of an improvement in time to mechanical ventilation or death.

Given the results of studies outlined above, remdesivir (RDV) in combination with TCZ could provide better efficacy, offering the potential to treat COVID-19 pneumonia in hospitalized populations more effectively than RDV alone. Extensive safety data have previously been generated on the use of TCZ in other indications. In addition, data for RDV and TCZ in patients with severe COVID-19 pneumonia have shown both agents to be well tolerated in this patient population. Therefore, a study investigating the safety and efficacy of combination therapy with RDV plus TCZ compared with RDV alone 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 WA42511 is a Phase III, randomized, double-blind, placebo-controlled, multicenter study to assess the efficacy and safety of TCZ in combination with RDV compared with matching placebo (PBO) in combination with RDV in hospitalized adult patients with severe COVID-19 pneumonia. The Sponsor intends to enroll approximately 650 patients who are diagnosed with COVID-19 pneumonia and meet the entry criteria in centers globally.

Patients must be at least 12 years of age with confirmed SARS-CoV-2 (COVID-19*)* infection per World Health Organisation (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 require $>$6 L/min supplemental oxygen to maintain oxygen saturation (SpO_2) $>$ 93% despite being on SOC, which may include, 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 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+RDV or a matching placebo+RDV, respectively. Study treatment will be given in combination with standard supportive care. The randomization will be stratified by geographic region (North America, Europe, and other) and a 2-level factor based on the assessment of the 7-category ordinal scale of clinical status at screening, with levels 4-5 and 6. "North America" includes the United States of America, "Europe" includes Spain and "Other" includes Russia and Brazil. The proportion of patients at level 6 of the ordinal scale will be capped at 25% of the overall study population.

Patients assigned to the TCZ+RDV arm will receive one infusion of TCZ 8 mg/kg after randomization, with a maximum dose of 800 mg, and patients assigned to the RDV+PBO arm will receive one infusion of PBO. Both arms will receive 200 mg RDV as the loading dose and 100 mg daily for the next 9 days in addition to SOC. Patients can have up to 2 doses of RDV prior to randomization.

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.

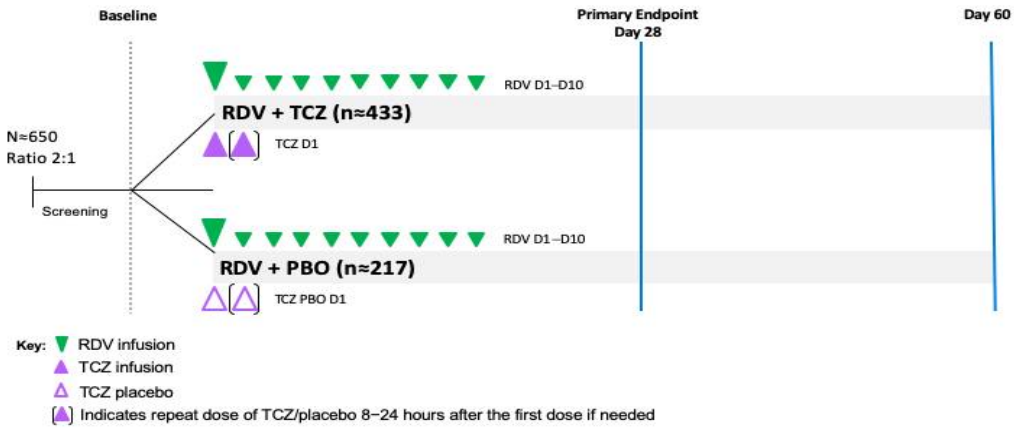
Remdesivir dosing is adjusted for patients that enter the study having received prior RDV. Patients who received RDV prior to randomization will not exceed 10 days of dosing in total (including RDV received prior to the study and during the study).

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 (AEs), 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 WA42511 Schema



n= number of patients; PBO= placebo; RDV=remdesivir; 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, and safety of TCZ+RDV compared with a matching TCZ placebo+RDV 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 the RDV plus TCZ arm compared with RDV plus PBO arm using the following endpoint:

- Time from randomization to hospital discharge or “ready for discharge” up to Day 28

  Hospital discharge or “ready for discharge” is defined as an ordinal score of 1 on the 7-category ordinal scale.

2.2.2 Secondary Efficacy Endpoints

- Time to mechanical ventilation or death up to Day 28, defined as the time from randomization to the first occurrence of mechanical ventilation or death (whichever occurs first)
- Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical status up to Day 28

- Clinical status as assessed by the investigator using a 7-category ordinal scale of clinical status on Days 7, 14, 21, 28 and Day 60 as follow-up.
- Proportion of patients requiring initiation of mechanical ventilation postbaseline up to Day 28 and up to Day 60 (patients who do not require mechanical ventilation at baseline)
- Proportion of patients who are alive and free of respiratory failure (patients requiring mechanical ventilation at baseline) at Day 28 and Day 60
- Duration of mechanical ventilation (patients who require mechanical ventilation at baseline) up to Day 28
- Time to Death up to Day 28 and up to Day 60
- Mortality on Days 14, 28, and 60 (proportions at specified time points)
- Time to recovery up to Day 28, defined as time from randomization to the time when a category of 2 on the 7-category ordinal scale (non-ICU hospital ward or "ready for hospital ward") not requiring supplemental oxygen, or better is observed
- Proportion of patients discharged or "ready for discharge" up to Day 28
- Proportion of patients who require initiation of mechanical ventilation postbaseline or die up to Day 28.

A pre-defined meta-analysis may be performed on mortality as well as the primary and selected secondary endpoints with Studies WA42380 and ML42528 (Section 4.5.4.5).

2.2.3 Exploratory Efficacy Endpoints

- Time to clinical improvement (TTCI) up to Day 28, defined as time from randomization to National Early Warning Score 2 (NEWS2) score of ≤ 2 maintained for 24 hours:

  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).

- Time to clinical failure up to Day 28, defined as the time from randomization to the first occurrence of death, mechanical ventilation, ICU admission, or withdrawal from study prior to discharge (whichever occurs first).

  For patients entering the study already on mechanical ventilation, clinical failure is defined as a one-category worsening on the ordinal scale, withdrawal from study prior to discharge, or death.

- Proportion of patients requiring initiation of ICU care post-baseline up to Day 28 and Day 60
- Duration of ICU stay up to Day 28
- Duration of supplemental oxygen use up to Day 60
- Proportion of patients requiring initiation of vasopressor use post-baseline up to Day 28

- Duration of vasopressor use up to Day 28
- Proportion of patients requiring initiation of extracorporeal membrane oxygenation (ECMO) postbaseline up to Day 28 and up to Day 60
- Duration of ECMO up to Day 28
- Organ failure-free days from randomization to Day 28
- SARS-CoV-2 viral load up to Day 28 and Day 60
- Time to reverse-transcriptase polymerase chain reaction (RT-PCR) SARS-CoV-2 negativity up to Day 60 (swab samples) and Day 14 (serum samples)

2.2.4 Biomarkers

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 biomarker analysis plan (BAP).

2.2.5 Safety Endpoints

- Incidence and severity of AEs, with severity determined according to National Cancer Institute Common Terminology Criteria for AEs (NCI CTCAE) v5.0
- The proportion of patients with any post-treatment infection up to Day 28 and Day 60
- Change from baseline in targeted clinical laboratory test results

2.3 DETERMINATION OF SAMPLE SIZE

The primary endpoint, time to discharge or "ready for discharge", is event driven. Based on the severe cohort receiving 10 days of RDV in Gilead's SIMPLE trial (Study GS-US-540-5773), the median time to discharge or "ready for discharge" was 11 days. Assuming a median time to discharge or "ready for discharge" of 11 days in the PBO+RDV arm, a hazard ratio of 1.3 or an approximately 2.5-day reduction in median time for TCZ+RDV vs PBO+RDV, and a 2:1 randomization to TCZ+RDV or PBO+RDV, approximately 650 patients are needed to accrue approximately 520 events to achieve approximately 80% power. Further sample size adjustments may be considered during the study based on external information. The sample size may be increased up to a maximum of approximately 800 randomized patients if fewer events than expected are observed or further shifts in SOC warrant reassessing sample size assumptions.

2.4 ANALYSIS TIMING

There may be up to three optional interim analyses. Should an interim occur before the end of the study, the SAP will be amended to account for the interim. For additional information about interim analyses, refer to Section 4.7.

If efficacy is declared based on an interim analysis, 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.

There will be an additional analysis on the final data when all patients have either reached Day 60 or withdrawn. A Clinical Study Report based on the final analyses from the study will be produced.

3. STUDY CONDUCT

The plan is to enroll approximately 650 or more patients, up to a total of 800, who 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 PBO, respectively. Both arms will be administered RDV. For both arms, if the clinical signs or symptoms worsen or do not improve, one additional infusion of blinded treatment of TCZ or PBO can be given, 8–24 hours after the initial infusion. Patients will be followed up for a total of 60 days after randomization.

If patients are discharged from hospital prior to Day 28, follow-up visits should be conducted on Day 14, Day 21, and Day 28. Follow-up visits on Day 14 and Day 21 may be conducted as telephone visits. Patients should return to the site for the Day 28-visit, if at all possible. 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 for discharged patients. Patients should return to the site for the Day 60 visit, if at all possible.

3.1 RANDOMIZATION, STRATIFICATION AND BLINDING

Patients will be randomized at a 2:1 ratio to receive blinded treatment of either TCZ+RDV or a matching PBO+RDV, respectively. The randomization will be stratified by geographic region (North America, Europe, and other) and a 2-level factor based on the assessment of the 7-category ordinal scale of clinical status at screening, with levels 4–5 and 6; and will occur through use of a permuted-block randomization method to ensure a balanced assignment to each treatment arm. The proportion of patients in the ordinal score 6 stratum will be capped at no more than 25% of the overall study population.

Study site personnel and patients will be blinded to TCZ treatment assignment during the study. The Sponsor and its agents will also be blinded to TCZ 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 DMC members and support staff as specified in the DMC Charter who may be Roche employees but independent of the study team.

If unblinding is necessary for a medical emergency (e.g., in the case of a serious AE [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, AEs of special interest (AESI), and any events requiring the review of aggregate unblinded data during the study, such as interim analyses.

The DMC will also evaluate safety according to policies and procedures detailed in a DMC Charter. Regular safety reviews will begin after approximately 45 patients (30 TCZ+RDV, 15 PBO+RDV) have been enrolled and reached 14-day follow-up. Early stopping criteria will be detailed in the DMC charter and a separate interim SAP (iSAP) should an interim occur. Further details of any efficacy interims are provided in Section 4.7. Interim analyses will be conducted by a statistical programmer and statistician independent from the study management team (SMT) and will be reviewed by the DMC. Interactions between the DMC and Sponsor will be carried out as specified in the DMC Charter.

The DMC initially consisted of Sponsor representatives not involved in any operational aspects of the study and a scientific oversight committee of external experts (responsibilities and operating principles of the DMC are described in a charter, the Internal Monitoring Committee and Scientific Oversight Committee Agreement). The DMC responsibilities transitioned to a fully independent data monitoring committee (iDMC) prior to the 3rd scheduled safety review after approximately 300 patients reached Day 28 of follow-up. The iDMC will also conduct a 4th scheduled data review

when at least 450 patients have reached Day 28. Data processing will be handled by a Sponsor statistician and statistical programmer independent of the SMT. Details are reported in the iDMC charter.

4. STATISTICAL METHODS

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

In all safety analyses, patients will be grouped according to the treatment 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). Efficacy analyses will be based on the mITT population, if not otherwise specified. Analysis of safety 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 who received any amount of TCZ or PBO, with patients grouped according to the treatment assignment at randomization (TCZ+RDV or PBO+RDV).

4.1.2 Safety Population

The safety population will consist of all patients who receive any amount of study medication (RDV and/or TCZ/PBO). In all safety analyses, patients will be grouped according to the treatment that the patients received rather than the treatment assigned at randomization (TCZ+RDV or PBO+RDV). If a patient only received RDV and did not receive TCZ/PBO, they will be included in the PBO + RDV arm. If a patient only received TCZ/PBO and did not receive study RDV, they will be included in the TCZ+RDV or PBO+RDV arms based on the TCZ/PBO treatment received.

4.2 ANALYSIS OF STUDY CONDUCT

The number of patients randomized, number of patients who complete the study, and number of patients discontinued early will be summarized. Reasons for premature study discontinuation will be listed and summarized. A listing of patients who discontinued early by treatment group will be produced through Day 28, the listing will include ordinal scale category at baseline and at last recorded visit for each patient, as well as discharge date and study day if the patient discontinued post discharge. The number of patients lost to follow-up and the number of patients who discontinued will be summarized by treatment group and whether discontinuation or loss to follow-up was before or after discharge.

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

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 along with a listing of the randomized patients by investigator.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic and baseline characteristics (including, but not limited to, age, sex, race, geographic region, ordinal scale category for clinical status, interleukin-6 [IL-6], mechanical ventilation, RDV prior to randomization, corticosteroids at baseline) will be summarized using means, standard deviations, medians and range 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
- Race
- Ethnicity
- Geographic region

4.3.2 Baseline Characteristics

- RDV prior to randomization
- Corticosteroid use at baseline
- Smoking history (Former/current user)
- Diabetes (Yes/ No)
- Heart disease (Yes/No)
- Hypertension (Yes/No)
- Weight
- NEWS2
- Ordinal scale category for clinical status
- IL-6
- CRP
- Ferritin
- D-Dimer

- Mechanical ventilation
- 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
- Number of days from last positive PCR or COVID-19 diagnosis at baseline
- Number of days of mechanical ventilation prior to randomization for patient in category 5 or 6 at baseline.

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 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, where Study Day 1 is the day of randomization. There will be a summary of all concomitant treatments, including those that were initiated prior to Study Day 1. Summaries of previous and concomitant steroid use will be produced. In addition, there will be summaries of previous and concomitant treatments given for COVID-19.

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 includes death, discharge from hospital and ICU admission and discharge, which will be summarized weekly in descriptive summaries, following the time windowing approach described below.

Deaths will also be captured on the 7-category 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 1).

Table 1 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 day of randomization.

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 discontinuation visits and any unscheduled visits. Time windows prior to Day 28, will be continuous from the midpoint between two consecutive study visits to the next midpoint, as in Table 2, except as indicated, and will be dependent on the schedule of assessments for each variable independently. An example of time windowing for the hematology parameters is shown below.

Table 2 Time Windows for Assigning Assessment Study Days to Study Visits for Hematology Parameters

Scheduled Study Day	[a]Efficacy Time Window
1 (Baseline)	≤ 1
2	2
3	3
5	> 3 to ≤ 5
7	>5 to ≤ 8
10	> 8 to ≤ 11
14	> 11 to ≤ 17
21	> 17 to ≤ 24
28	> 24 to ≤ 28
35	> 28 to ≤ 38
45	> 38 to ≤ 52
60	> 52 to ≤ 67

[a] From Week 2 onwards use value nearest to scheduled Study day.

Where there is more than one efficacy assessment within a time window, then the non-missing assessment nearest to the scheduled time point 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.

For patients who receive study drug on Day 1, baseline is defined at the last pre-dose assessment before administration of either study drug on or prior to Day 1. For patients who do not receive study drug on Day 1, baseline is the last assessment on or prior to Day 1. Where baseline data is not available for the ordinal scale or biomarkers, the first available assessment (up to Day 2) will be used as baseline.

4.5 EFFICACY ANALYSIS

All efficacy analyses will use the mITT population unless otherwise stated.

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 4.5.6.

The primary analysis will occur after the last patient has reached Day 28, analyses and censoring rules will be updated for analyses at or up to Day 60 when the database is refreshed.

Where the stratification factor of baseline ordinal scale (4-5, 6) is included in an analysis, this will take the value at randomization and may differ from the summary of baseline ordinal scale categories included in the baseline characteristics, which uses post-randomization data.

4.5.1 Primary Efficacy Endpoint

The primary efficacy objective for this study is to evaluate the efficacy of the TCZ+RDV arm compared with the PBO+RDV arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

> The primary efficacy endpoint is the time to hospital discharge or "ready for discharge". Hospital discharge or "ready for discharge" is defined as a score of 1 on the 7-category ordinal scale.

The other primary estimand attributes are:

- Population: Patients with severe COVID-19 pneumonia as per the inclusion/exclusion criteria specified in the protocol (mITT)
- Treatments: TCZ + RDV versus PBO + RDV
- Intercurrent events:
    - Events that lead to study withdrawal or loss to follow-up: Hypothetical strategy, i.e., patients will be censored at time of their last ordinal scale assessment, unless they die on or prior to Day 28.
- Summary measures:
    a) The hazard ratio (95% confidence interval [CI])
    b) Kaplan-Meier plot
    c) Cumulative incidence function (CIF), i.e., the cumulative probability of being discharged or "ready to discharge" over the 28 day follow-up period. The CIF of the competing event of death prior to discharge will also be summarized.
    d) Median event time in each treatment arm (95% CI)

Patients will meet the endpoint at the time of discharge or the time that they achieve category 1 of the 7-category ordinal scale, whichever occurs first, provided that:

- they do not have any further ordinal scale assessments > category 1 on or prior to Day 28
- they are not re-hospitalized on or prior to Day 28
- they do not die on or prior to Day 28

Patients who do not meet the event at the point of discharge or category 1 due to re-hospitalization or ordinal scale assessments > category 1 will still be eligible to meet the event at a later time provided the above conditions are met. Patients who die by Day 28, regardless of discharge and ordinal scale category prior to death, and patients who remain hospitalized at Day 28 with an ordinal scale category >1 will not be considered as having met the endpoint.

The time to hospital discharge or "ready for discharge" will be compared between the TCZ+RDV and the PBO+RDV arms using the stratified log-rank test with region (North America, Europe, Other) and baseline ordinal score (4-5, 6) included as the stratification factors at Day 28 using the mITT population. This will be tested at a two-sided 5% significance level.

The treatment groups will be compared descriptively using a Cox proportional hazards model adjusted for the stratification factors of baseline ordinal score (4-5 or 6) and

region (North America, Europe or Other). Kaplan-Meier plots will be produced and CIF plots will be produced for both the primary outcome and the competing risk of mortality.

Intercurrent events, such as events leading to loss to follow-up or discontinuation for any reason prior to achieving the event or patients who do not have the event, will be accounted for through rules, as described in Table 3 below.

Table 3 Time to Hospital Discharge or "Ready for Discharge" Censoring Rules

Event	Censor	Date and Time
Death	Yes	Day 28
Withdrawal or lost to follow-up for any reason prior to discharge or "ready for discharge" criterion met (no event or death recorded including post-withdrawal or lost to follow-up)	Yes	last recorded ordinal scale assessment
Not discharged or "ready for discharge"	Yes	Day 28

In addition, a summary of hospital discharges over time will be produced.

4.5.2 Controlling for Type I Error

In the event of an interim analysis, the primary endpoint will be adjusted based on the alpha spending function detailed in an iSAP.

Secondary to time to discharge or "ready for discharge", 3 key secondary endpoints will be tested in a simple gated hierarchy starting with the primary endpoint. The hierarchy will be:

1. Time from randomization to hospital discharge or "ready for discharge" up to Day 28
2. Time to mechanical ventilation or death up to Day 28
3. Clinical status as assessed by the investigator using a 7-category ordinal scale of clinical status on Day 14
4. Time to death up to Day 28

Each endpoint will be tested with a fixed two-sided 0.05 error rate if the previous endpoint reaches significance starting with the primary. Since each successive hypothesis will be gated on the previous test sequentially, there is no penalty on the type I error (Dmitrienko and Tamhane 2010).

All other secondary endpoints defined in the protocol have clinical relevance in understanding the therapeutic benefit of TCZ. A treatment effect may be observed that may not meet statistical significance, but may still be considered clinically meaningful. Therefore, all other secondary endpoints will be tested without adjustment.

4.5.3 Type I Error Controlled Secondary Endpoints

4.5.3.1 Time to Mechanical Ventilation or Death up to Day 28

This secondary efficacy objective evaluates the efficacy of the TCZ+RDV arm compared with the PBO+RDV arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

> The secondary efficacy endpoint time to mechanical ventilation or death up to Day 28 is the time from randomization to the first occurrence of mechanical ventilation or death (whichever occurs first). For patients already on mechanical ventilation at baseline, only death will be counted as an event. If no event occurs within the time frame, the patient is censored on Day 28.

Time to mechanical ventilation or death will be compared between the TCZ+RDV and the PBO+RDV groups using the stratified log-rank test, with geographic region (North America, Europe, and Other) and baseline ordinal scale (4-5, 6), included as the stratification factors using the mITT population. A Kaplan-Meier plot and median time to event with 95% CIs will also be presented. In addition, the treatment groups will be compared descriptively using a Cox proportional hazards model adjusting for the stratification factors applied at randomization and a hazard ratio with a 95% CI will be produced.

The other estimand attributes are:

- Population: Patients with severe COVID-19 pneumonia as per the inclusion/exclusion criteria specified in the protocol (mITT)
- Treatments: TCZ+RDV versus PBO+RDV
- Intercurrent events:
    - Study withdrawal or loss to follow-up: The Sponsor will use public records in accordance with local regulations to identify deaths even after loss to follow-up. If no such events are found, a hypothetical strategy is used, i.e., patients who withdraw or are lost to follow up prior to discharge will be censored at time of withdrawal or last recorded visit.
    - Treatment is modified or withdrawn: Treatment policy strategy, i.e. this intercurrent event is ignored for this endpoint.
    - Patient discharged from hospital: Treatment policy strategy, i.e. this intercurrent event is ignored for this endpoint.
- Population summary measures:
    a) Median time to event in each treatment arm with 95% confidence intervals
    b) The hazard ratio (95% confidence interval [CI])
    c) Kaplan-Meier plot

Table 4 Time to Mechanical Ventilation or Death Censoring Rules

Event	Censor	Date and Time
Withdrawal or lost to follow-up for any reason prior to discharge prior to criterion met (not followed by death)	Yes	last recorded vital signs assessment
Withdrawal or lost to follow-up for any reason on or after the day of discharge prior to criterion met (not followed by death or readmittance)	Yes	Day 28
Criterion not met	Yes	Day 28

4.5.3.2 Clinical Status on Day 14

This secondary efficacy objective evaluates the efficacy of the TCZ+RDV arm compared with the PBO+RDV arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

> The secondary efficacy endpoint clinical status as assessed by the investigator using a 7-category ordinal scale of clinical status. The latest available assessment up to Day 14 is used.

Assessment of patient status using an ordinal scale will be recorded at baseline and once daily in the morning (between 8:00 a.m. and 12:00 p.m.) 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 ≤2 L 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

For patients who withdraw before the assessment time, their last postbaseline ordinal category prior to withdrawal will be used in the analysis, unless death within the time frame was captured from public records or otherwise; in which case death will be used in the analysis. A death or discharge (unless the patient is re-admitted within 12 hrs) will always be carried forward to all subsequent assessments regardless of what is recorded for the ordinal scale. In addition to imputing the ordinal scale at Day 14 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. If a patient is re-admitted then the ordinal scale data from the point of re-admittance will be used.

Clinical status will be compared between the TCZ+RDV and PBO+RDV arms using a proportional odds model accounting for stratification factors at randomization in the model (region [North America, Europe, Other] and baseline ordinal scale [4-5, 6]) using the mITT population as the main analysis. The odds ratio, p-value, and 95% confidence interval will be presented.

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

In addition, the difference in distributions between treatment arms will be evaluated using a non-parametric method, the Van Elteren test, including the stratification factors at randomization (region [North America, Europe, Other] and baseline ordinal scale [4-5, 6]). 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.

Clinical status will also be assessed using a linear regression approach with Huber-White sandwich estimates for the standard errors. The mean and associated 95% CI in each treatment arm will be presented along with the difference in means and a 95% CI for the difference.

The ordinal scale will be summarized by treatment arm showing the number and percentage in each category. Missing data, defined as patients who have not died or been discharged and have no subsequent measurements, will be entered into its own category. Deaths and discharge will be carried forward as above. 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. The summary and stacked bar charts will also be produced using the data after the last postbaseline ordinal category has been carried forward.

The other estimand attributes are:

- Population: Patients with severe COVID-19 pneumonia as per the inclusion/exclusion criteria specified in the protocol (mITT)
- Treatments: TCZ+RDV versus PBO+RDV
- Intercurrent events:
    - Discharge from hospital: Composite strategy, this is considered a value of 1. This assessment of 1 is carried forward to all subsequent

assessments, unless a death is observed or the patient is re-hospitalized until Day 14.

- Death: Composite strategy, this is considered a value of 7.
- Study withdrawal or loss to follow-up: For patients who withdraw before the assessment time, a hypothetical approach is used, i.e. their last post-baseline ordinal category prior to withdrawal will be used in the analysis. If a death event up to Day 14 is known (e.g. from public records), a value of 7 ("Death") will be used instead.
- Treatment is modified or withdrawn: Treatment policy strategy, i.e. this intercurrent event is ignored for this endpoint.

– Population summary measures:

a) The frequency distribution of clinical status by arm will be tabulated and shown graphically.

b) The odds ratio (95% CI)

4.5.3.3 Time to Death up to Day 28

This secondary efficacy objective evaluates the efficacy of the TCZ+RDV arm compared with the PBO+RDV arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

The secondary efficacy endpoint time to death up to Day 28 is the time from randomization to death from any cause. If no event occurred within the time frame, the patient is censored on Day 28.

Time to death will be compared between the TCZ+RDV and the PBO+RDV groups using the stratified log-rank test, with geographic region (North America, Europe, and Other) and baseline ordinal scale (4-5, 6), included as the stratification factors using the mITT population. A Kaplan-Meier plot and median time to event with 95% CIs will also be presented for each treatment arm. In addition, the treatment groups will be compared descriptively using a Cox proportional hazards model adjusting for the stratification factors applied at randomization and a hazard ratio with a 95% CI will be produced.

The other estimand attributes are:

– Population: Patients with severe COVID-19 pneumonia as per the inclusion/exclusion criteria specified in the protocol (mITT)

– Treatments: TCZ+RDV versus PBO+RDV

– Intercurrent events:

- Study withdrawal or loss to follow-up: The Sponsor will use public records in accordance with local regulations to identify deaths even after loss to follow-up. If no such events are found, a hypothetical strategy is used, i.e., patients who withdraw or are lost to follow up prior to discharge will be censored at last known alive date.

- Treatment is modified or withdrawn: Treatment policy strategy, i.e. this intercurrent event is ignored for this endpoint.
- Patient discharged from hospital: Treatment policy strategy, i.e. this intercurrent event is ignored for this endpoint.

– Population summary measures:

a) Median time to event in each treatment arm with 95% confidence intervals

b) The hazard ratio (95% confidence interval [CI])

c) Kaplan-Meier plot

Table 5 Time to Death Censoring Rules

Event	Censor	Date and Time
Withdrawal or lost to follow-up for any reason prior to discharge prior to criterion met (not followed by death)	Yes	last known alive date
Withdrawal or lost to follow-up for any reason on or after the day of discharge prior to criterion met (not followed by death or readmittance)	Yes	Day 28
Criterion not met	Yes	Day 28

4.5.4 Other Secondary Endpoints

4.5.4.1 Time to Event Analyses

Other secondary time to event endpoints will be compared between the TCZ+RDV and the PBO+RDV groups using the stratified log-rank test, with geographic region (North America, Europe, and Other) and baseline ordinal scale (4-5, 6), included as the stratification factors using the mITT population. A Kaplan-Meier plot and median time to event in each treatment arm with 95% CIs will also 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 and a hazard ratio with a 95% CI will be produced. For endpoints representing an improvement, a cumulative incidence function plot for the event of interest and mortality will be produced.

Other secondary time to event endpoints include:

- Time to Death up to Day 60.
  - This endpoint is defined identically to Time to Death up to Day 28 described in Section 4.5.3.3, except that patients are censored at Day 60 instead of Day 28 if no event occurred. Analyses, handling of intercurrent events and population summary measures are as described previously.
- Time to improvement in ordinal clinical status up to Day 28 (days)
  - Defined as time from randomization to the time when at least a 2-category improvement in the 7-category ordinal scale is observed

- Time from randomization to recovery up to Day 28 (days)
  - Defined as time from randomization to the time when a clinical status of discharged or “ready for discharge” (ordinal scale category 1), or non-ICU hospital ward or “ready for hospital ward” not requiring supplemental oxygen (ordinal scale category 2) is reached

The estimand for all time to event data is the difference in distributions between TCZ+RDV and PBO+RDV using the stratified log-rank test. In case no events occurred up to Day 28, patients are censored on Day 28.

Intercurrent events are those that occur after randomization and either preclude observation of the variable or affect its interpretation. Intercurrent events, such as events leading to loss to follow-up or discontinuation for any reason prior to achieving the event or patients who do not have the event, will be accounted for through censoring rules as shown in the tables for each endpoint.

- Time to improvement in ordinal clinical status up to Day 28 (days)

  Defined as time from randomization to the time when at least a 2-category improvement in the 7-category ordinal scale is observed. For patients that are discharged and the ordinal scale assessment has not been completed at discharge, they will be assumed to be in Category 1 of the ordinal scale at the point of discharge, unless they are re-admitted within 12 hours. Intercurrent events, such as events leading to loss to follow-up, or discontinuation from the study for any reason prior to achieving the event or patients who do not have the event, will be accounted for through censoring rules, as described in the table below. Deaths will be censored at Day 28.

Table 6 Time to Improvement in Ordinal Clinical Status Censoring Rules

Event	Censor	Date and Time
Death (regardless of criterion met)	Yes	Day 28
Withdrawal or lost to follow-up for any reason prior to criterion met	Yes	Last recorded ordinal scale assessment
Criterion not met	Yes	Day 28

- Time from randomization to recovery up to Day 28 (days).

  Recovery is defined as clinical status discharged or “ready for discharge” (ordinal scale category 1), or non-ICU hospital ward or “ready for hospital ward” not requiring supplemental oxygen (ordinal scale category 2).

  Patients will meet the endpoint at the time that they achieve category 2 or better on the 7-category ordinal scale or at the time that they are discharged, provided that they do not have any further ordinal scale assessments > category 2 on or prior to Day 28. Patients who die by Day 28, regardless of discharge and ordinal scale category prior to death, and patients who remain hospitalized at Day 28 with an ordinal scale category >2 will not be considered as having met the endpoint. Intercurrent events, such as events leading to loss to follow-up, or discontinuation for any reason prior to achieving the event or patients who do not have the event, will be accounted for through censoring rules, as described in the table below. Deaths will be censored at Day 28.

Table 7 Time to Recovery Censoring Rules

Event	Censor	Date and Time
Death (regardless of criterion met)	Yes	Day 28
Withdrawal or lost to follow-up for any reason prior to criterion met	Yes	Last recorded ordinal scale assessment
Criterion not met	Yes	Day 28

4.5.4.2 Incidence Endpoints

Secondary efficacy incidence endpoints will be analyzed using the Cochran-Mantel-Haenszel test statistic (Mantel 1963), adjusted by the stratification factors at baseline geographic region (North America, Europe, and Other) and 7 point ordinal score at randomization (4-5, 6) 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, along with the proportions in each treatment arm and associated 95% CIs.

- Proportion of patients requiring initiation of mechanical ventilation postbaseline up to Day 28 and up to Day 60

  For patients who are not on mechanical ventilation at baseline (ordinal score 4 or less) the incidence of mechanical ventilation by Day 28 and Day 60 will be identified from the Vital Signs and Oxygen Saturation eCRF. For patients who have withdrawn or died prior to Day 28 or Day 60, the non-responder rule will be applied, i.e., it will be assumed that the patient required mechanical ventilation by Day 28 or Day 60 in the analysis. Patients without mechanical ventilation prior to discharge, will be assumed to be responders in the analysis except for the following cases:

  1. The patient is readmitted to hospital within 12 hours. The patient will be treated as having not been discharged, and any initiation of mechanical ventilation will be counted
  2. The patient dies by Day 28 or Day 60, which will be counted as an event.

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

- Proportion of patients who are alive and free of respiratory failure (patients requiring mechanical ventilation at baseline) at Day 28 and Day 60

  The incidence of patients who come off ventilator support, for patients who enter the study on mechanical ventilation will be derived from the Vital Signs and Oxygen Saturation eCRF. Patients who transition to oxygen support other than mechanical ventilation or ECMO will be counted as extubated, unless they die prior to Day 28 or Day 60 in the respective analyses in which case they will be counted with patients who remain on mechanical ventilation. Patients who are re-intubated will be counted as on mechanical ventilation unless they come off ventilation again before the end of the 28-day follow-up. Patients who die up to Day 28 or Day 60 in the respective analyses will be counted as non-responders and counted with patients who continue on mechanical ventilation. Patients who withdraw from the study who are not yet extubated, prior to Day 28 or Day 60 in the respective analyses will be counted as remaining on mechanical ventilation.

- Difference in mortality at Day 14, 28, and Day 60

  The difference in proportion of patients that have died will be compared at the specified time points and will be summarized by the proportion who died in each group and the difference in proportions as well as 95% CIs for each estimate. All deaths post discontinuation and discharge will be included in this analysis as long as the death did not occur after patient consent was withdrawn unless information can be sourced from public records. Patients who withdraw prior to discharge will be imputed as having survived. Deaths occurring between each visit, and cumulative deaths by visit will be summarized descriptively to Day 60.

- Proportion of patients discharged or “ready for discharge” up to Day 28

The proportion of patients who meet the primary endpoint of Time to Hospital Discharge or “Ready for Discharge” up to Day 28 defined in Section 4.5.1 will be compared across treatment arms.

- Proportion of patients who require initiation of mechanical ventilation postbaseline or die up to Day 28

  The proportion of patients who meet the secondary endpoint of Time to Mechanical Ventilation or Death up to Day 28 defined in Section 4.5.3.1 will be compared across treatment arms.

4.5.4.3 Ordinal endpoints

- Clinical status as assessed by the investigator using a 7-category ordinal scale of clinical status on Day 7, 21, 28, and Day 60 as follow-up.

  This endpoint is defined in the same way as Clinical Status on Day 14 described in Section 4.5.3.2. Analyses and summaries are as described previously.

4.5.4.4 Duration endpoints

- Duration of mechanical ventilation up to Day 28

Duration of mechanical ventilation (in patients who require mechanical ventilation at baseline) is defined as the number of days postbaseline to Day 28 that a patient is alive and requiring mechanical ventilation. Duration of mechanical ventilation will be derived from the Vital Signs and Oxygen Saturation eCRF; if mechanical ventilation or ECMO is recorded for any part of the day, the day will be counted. Duration of mechanical ventilation will be 28 days if the patient is mechanically ventilated from Day 1 to Day 28. Duration of mechanical ventilation will be coded as 28 days if a patient dies on or prior to Day 28.

For patients withdrawn early from the study but not discharged, if patients were on mechanical ventilation at the point of discontinuation it will be assumed that the remainder of days to Day 28, are mechanical ventilation days. For patients not requiring mechanical ventilation at the point of withdrawal, it will be assumed the period from withdrawal to Day 28 are not mechanical ventilation days. For patients who are discharged, days from discharge to Day 28 will not be counted as mechanical ventilation days.

Duration of mechanical ventilation will be analyzed using a linear regression approach with Huber-White sandwich estimates for the standard errors, including the stratification factors at randomization (region [North America, Europe, Other] and ordinal scale category at baseline [4-5, 6]) as covariates. The mean duration of mechanical ventilation for each treatment group and the corresponding 95% CIs for the means will be presented along with the P-value, the difference in means and a 95% CI for the difference.

Duration of mechanical ventilation will also be summarized descriptively using the means, along with 95% CIs, by treatment group for those patients alive at Day 28, with a count of the number of patients who died and are counted as having 28 days duration of ventilation.

4.5.4.5 Meta-Analysis

The details of the meta-analysis and data handling from multiple trials will be included in a separate meta-analysis plan.

4.5.5 Exploratory Efficacy Endpoints

- TTCI up to Day 28

  Defined as time from randomization to NEWS2 of ≤ 2 maintained for 24 hours. 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. No imputation will be done for missing components of the NEWS2 score and only complete assessments will be considered in the analysis.

  Patients who have a score of ≤ 2 at baseline will be analyzed 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. Patients who die will be censored at Day 28. Patients who withdraw or are lost to follow up prior to meeting the criterion will be censored at their last vital sign assessment. In case no events occurred up to Day 28, patients are censored on Day 28.

  This endpoint will be analysed in a similar way to the time to event endpoints in Section 4.5.4.1.

- Time to clinical failure up to Day 28 including ICU admission

  Defined as the time from randomization to the first occurrence of death, mechanical ventilation, ICU admission, or withdrawal from the study prior to discharge (whichever occurs first). For patients entering the study already on mechanical ventilation, clinical failure is defined as a one-category worsening on the ordinal scale, withdrawal from study or death. Withdrawals after hospital discharge will not be considered as having met clinical failure, unless a later death is recorded in which case they will have met the event upon death. Intercurrent events, such as events leading to loss to follow-up or discontinuation for any reason prior to the event or patients who do not have the event, will be accounted for through

censoring rules as described in the table below.Clinical failure status as defined above will be summarized descriptively by visit.

Table 8 Time to Clinical Failure Censoring Rules

Event	Censor	Date and Time
Lost to follow-up for any reason prior to discharge prior to criterion met (not followed by death)	Yes	last recorded vital signs assessment
Withdrawal or lost to follow-up for any reason on or after the day of discharge prior to criterion met (not followed by death or readmittance)	Yes	Day 28
Criterion not met	Yes	Day 28

This endpoint will be analysed in a similar way to the time to event endpoints in Section 4.5.4.1.

- Proportion of patients requiring initiation of ICU care postbaseline up to Day 28 and Day 60

  Proportion of patients admitted to the ICU after randomization will be tabulated by treatment with the number of patients and 95% CIs in each treatment arm, the difference in proportions, 95% CIs and the p-value. Patients who die prior to ICU admission in hospital will be treated as an ICU admission.

  This endpoint will be analysed in a similar way to the incidence endpoints in Section 4.5.4.2.

- Duration of supplemental O2 (days) to Day 60

  Duration of supplemental O2 (days) will also be derived from the vital signs and oxygen saturation log, where study days with any supplemental oxygen will be summed up to and including Day 60. Patients without any supplemental O2 use will assigned a duration of zero days. For missing data, the last observation postbaseline 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 Day 60 were on supplemental oxygen. For patients not using supplemental oxygen at point of withdrawal it will be assumed supplemental oxygen is not required to Day 60. For patients that are discharged, days from discharge to Day 60 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 and if there is no end date recorded it

will be assumed to have continued until Day 60). Duration of supplemental oxygen use will be 60 days if a patient dies on or prior to Day 60.

This endpoint will be analyzed in a similar way to the duration endpoints in Section 4.5.4.4.

In addition, the number and the proportion of patients on supplemental oxygen using the observed data will be summarized by visit to Day 60.

The proportion requiring supplemental O2 after discharge will also be summarized by treatment.

- Duration of ICU stay (days) up to Day 28

  Duration of ICU stay (days) will be calculated as the sum of the number days spent in ICU up to and including Day 28 based on the admission and discharge dates from the ICU stay information log; (ICU discharge date – ICU admission date + 1 day). Multiple periods of ICU stay will be summed. Patients without any ICU stays will be assigned a duration of zero days.

  For patients that are discharged, any ongoing ICU stays without an end date, their ICU end date will be imputed as date of discharge 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 Day 28 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 Day 28. For patients that die on or prior to Day 28 all days will count as being in the ICU.

  This endpoint will be analysed in a similar way to the duration endpoints in Section 4.5.4.4.

- Vasopressor and ECMO use (days) up to Day 28 and up to Day 60

  Incidence of vasopressor use (from concomitant medication records) and incidence of ECMO use postbaseline by Day 28 (and separately to Day 60) will be summarized descriptively.

  Duration of vasopressor use (days) and ECMO (days) postbaseline to Day 28 will be summarized using the mean along with 95% CIs for the mean by treatment group. ECMO is collected daily and the total 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 Day 28 will use the upper bound of the Day 28-time window as the end date for the duration. Duration of vasopressor use and ECMO postbaseline to Day 28 will also be summarized in the subgroup of patients with vasopressor use and the subgroup of patients with ECMO use, respectively.

- Organ failure-free days up to Day 28

Days without organ failure will be summarized descriptively through Day 28. 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 $\leq$ 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 $\geq$ 2 mg/dL, hepatic failure is defined as bilirubin $\geq$ 2 mg/dL and coagulation failure is defined as a platelet count of $\leq 80 \times 10^3/mm^3$. 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.

- SARS-CoV-2 viral load over time

  SARS-CoV-2 viral load over time will be summarized descriptively by time point and treatment group. Both the actual values and change from baseline will be summarized. The number and proportion of patients negative and positive will be displayed, and the quantitative result will be summarized. Area under curve (AUC) will be calculated using the trapezoidal method adjusted by the datetime of the last available assessment for each patient and summarized by treatment arm with mean (log[AUC]) and 95% CI's. The adjustment will be performed by dividing by the datetime of the last assessment. The AUC will only be calculated for patients with at least two virology assessments.

  Any values reported as BLQ will be set to the lower limit of quantitation for the assay minus 1. Any values reported as negative will be set to the half of lower limit of quantitation for the assay.

  This analysis will be conducted for swab samples and serum samples separately.

- Time to reverse-transcriptase polymerase chain reaction (RT-PCR) SARS-CoV-2 virus negativity

  Defined as days from randomization to when a negative Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR) SARS-CoV-2 virus assessment result is observed up to Day 60 for swab samples and up to Day 14 for serum samples. Time

to reverse-transcriptase polymerase chain reaction (RT-PCR) SARS-CoV-2 virus negativity will be compared descriptively using a Cox proportional hazards model adjusted for the stratification factors of baseline ordinal score (4-5 or 6) and region (North America, Europe or Other) in the subgroup of patients positive at baseline. A Kaplan-Meier plot and a CIF plot of the event of interest and mortality will be produced. Patients who die prior to achieving negativity will be censored at Day 60 in the swab analyses and Day 14 in the serum analyses. Patients who discontinue or are lost to follow up prior to meeting the criterion will be censored at their last virology assessment. In case no events occurred up to Day 60 in the swab analysis and Day 14 in the serum analysis, patients are censored on Day 60 and Day 14, respectively.

This analysis will be conducted for swab samples and serum samples separately.

4.5.6 Subgroup Analyses

Exploratory subgroup analysis may be performed either to supplement the topline results or as a sensitivity analysis. Analysis of biomarkers will be described in a separate BAP in which any biomarker analyses will be specified.

The primary endpoint will be analyzed by baseline ordinal score using a forest plot of the hazard ratios at each baseline level (except Category 7) including 95% CI's.

The following exploratory subgroups will be investigated for the primary endpoint and Time to Death up to Day 28 using forest plots:

- Baseline ordinal score
- Steroids (Y/N)
- Mechanical Ventilation (Y/N)
- Age (12-17, 18-64, 65 and up)
- Sex (M/F)
- Region (North America, South America, Europe [including Russia])
- Race/Ethnicity

4.5.7 Sensitivity and Tipping Point Analyses

4.5.7.1 Prognostic Covariates

Models adjusted for prognostic covariates such as age and individual baseline ordinal score category may be fit to the data for the primary and key secondary endpoints.

4.5.7.2 Clinical Status at Day 14

A multiple imputation approach for missing data will be performed using bootstrapping by sampling with replacement from non-missing data within treatment group and strata, assuming the data were missing at random. The imputed data will be combined with the non-missing data to give complete data sets of the same size as the mITT population. This will be repeated to obtain 100 complete data sets for each set of imputation

parameters. The ordinal logistic regression model (adjusted by stratification factors) will be fitted to each of the datasets to obtain the p-value and odds ratio for each of the complete datasets. The averages of the p-values and odds ratios from the 100 data sets will be presented.

To test the sensitivity of these results to the missing-at-random assumption for each of the 100 data sets, the imputed data points will then be adjusted independently by treatment group to worse or better outcomes according to a delta (addition of 1 to 6 or subtraction 1 to 6). After the delta is applied, adjusted scores below 1 or above 7 will be capped at 1 or 7, respectively, so that the data are within the range of possible values for the ordinal scale. For each delta adjustment, the ordinal logistic regression model (adjusted by stratification factors) will be fitted to each of the datasets to obtain the p-value and odds ratio for each of the complete datasets. The averages of the p-values and odds ratios from the 100 data sets will be presented for each of the delta adjustments.

The tipping point will be defined as the delta adjustments at which the p-value changes from statistically significant to not statistically significant (or vice versa).

4.5.7.3 Primary Time to Event Endpoint

A sensitivity analysis for the primary endpoint analysed in the mITT population will be conducted as described below.

In the following discussion, it is assumed that the mean time to event in the TCZ+RDV arm is greater than in the PBO+RDV arm. If, in reality, the mean time to event is greater in the PBO+RDV arm than in the TCZ+RDV arm, then the roles of the two groups in this section will be reversed.

The tipping point analysis will impute actual event times for patients whose observed time to event is censored due to events other than death.

Times to clinical response will not be imputed for patients whose time to event is:

- Observed
- Censored due to death
- Censored due to end of follow-up at Day 28

Step 1

For the primary endpoint, event times will be imputed by assigning a censored event time of 28 days to patients in the TCZ+RDV arm and an event time equal to the time of censoring for patients in the PBO+RDV arm. This is a “worst case” scenario. If the (non-) significance of the test for the main effect of treatment is unchanged in this

sensitivity analysis compared to the main efficacy analysis, no further imputation will take place.

For the time to mechanical ventilation or death and the time to death, event times will be imputed by assigning a censored event time of 28 days to patients in the PBO+RDV arm and an event time equal to the time of censoring for patients in the TCZ+RDV arm. This is a "worst case" scenario. If the (non-) significance of the test for the main effect of treatment is unchanged in this sensitivity analysis compared to the main efficacy analysis, no further imputation will take place.

Step 2

Multiple imputation will be used to impute times to event as follows, exploiting the lack of memory property of the exponential distribution.

Additional times to event, beyond those actually observed, will be drawn from an exponential distribution. The mean of the distribution will vary between treatment groups. Details of how the means will be calculated are given below. If the total time to event (that is the sum of the actual censored and additional times to event) is greater than 28 days, the time to event will be censored at Day 28. Otherwise, the time to event will be set to the largest integer less than or equal to the sum of the observed and imputed additional times.

The mean of the distribution of the additional time to event in the PBO+RDV arm will be set to the Kaplan-Meier estimate of the mean time to event based on the data observed in the study. The mean of the distribution of the additional time to event in the PBO+RDV arm will be defined as follows.

Write $\hat{\mu}_{PBO}$ and $\hat{\mu}_{TCZ}$ as the Kaplan-Meier estimates of the mean time to event in the PBO+RDV and TCZ+RDV arms respectively. Similarly, write $\mu^*_{PBO}$and μ^*_{TCZ} to represent the corresponding means of the imputed additional times to event.

Thus, $\mu^*_{PBO} = \hat{\mu}_{PBO}$ as defined above.

Define $\mu^*_{TCZ} = \mu^*_{PBO} - \rho(\hat{\mu}_{PBO} - \hat{\mu}_{TCZ})$for some real number ρ.

ρ therefore provides a scale on which the tipping point can be measured.

- ρ = 1 implies that the mean of the imputed additional time to event in the TCZ+RDV arm is equal to the observed mean.

- ρ = 0 implies that the mean imputed time is equal to that observed in the PBO+RDV arm (so that treatment with TCZ affords no benefit)

- As ρ → ∞, the benefit of TCZ increases because the mean of the additional time to event decreases and eventually all imputed additional times to event tend to zero.

- As ρ → -∞, the benefit of TCZ decreases because the mean of the additional time to event increases and eventually all imputed times to event are censored at Day 28.

The number of imputations performed for each value of ρ will be chosen so that the estimates derived from the tipping point analysis appear reasonably stable. The number of imputations required may differ for different values of ρ.

The distribution of the p-value associated with a given value of ρ will be summarized by its mean, median, 10th and 90th quantiles.

The tipping point will be defined as the value of ρ at which the mean p-value changes from statistically significant to not statistically significant (or vice versa).

The results of the tipping point analysis will be presented in graphical and tabular form.

Practical limits of the values of ρ

The mean of an exponential distribution must be greater than 0. This implies that

$$\hat{\mu}_{PBO} - \rho(\hat{\mu}_{PBO} - \hat{\mu}_{TCZ}) > 0$$

So that

$$\rho_{min} = \frac{\hat{\mu}_{PBO}}{(\hat{\mu}_{PBO} - \hat{\mu}_{TCZ})}$$

Where ρ_{min} is the minimum possible value of ρ.

Similarly, as ρ increases, the probability that an imputed additional time to event is censored at Day 28 increases. Thus, a reasonable upper limit for ρ, $\rho_{max}(\pi)$ say, is given by

$$p(imputed\ additional\ time\ to\ event > 28\ |\rho) > \pi$$

Or

$$exp(\frac{-28}{\hat{\mu}_{PBO} - \rho_{max}(\pi) \times (\hat{\mu}_{PBO} - \hat{\mu}_{TCZ})}) > \pi$$

So that

$$\rho_{max}(\pi) = \frac{\frac{28}{log(\pi)} + \hat{\mu}_{PBO}}{\hat{\mu}_{PBO} - \hat{\mu}_{TCZ}}$$

Since the imputed additional survival time is added to the censored survival times actually observed, values slightly less than ρ_{max} will also have little or no effect on the result of the analysis, so the value of ρ_{max} defined here is somewhat conservative.

4.5.7.4 Time to Death and Time to Mechanical Ventilation or Death up to Day 28

The main analyses of these endpoints censor patients who withdrawal or are lost to follow up on or after the day of discharge and before the criterion is met (not followed by death or re-hospitalization) at Day 28. The main analyses also censor patients who withdraw or are lost to follow up prior to discharge and before the criterion is met (not followed by death) on the last known alive date for the Time to Death analysis and the last recorded vital signs assessment for the Time to Mechanical Ventilation or Death analysis. To evaluate the effect of this approach, tipping point analyses, similar to those described above, will be conducted for these endpoints.

For each of these endpoints, two tipping point analyses will be conducted. In the first, additional times to event will be imputed for only for patients who withdraw or are lost to follow up following discharge. In the second, additional times to event will be imputed for all patients who withdraw or are lost to follow up.

Mean additional times to event will be generated as described in Section 4.5.7.3 and will be based on the times to event observed in each group. ρ will be defined so that a positive value indicates a benefit associated with TCZ.

4.6 SAFETY ANALYSES

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

4.6.1 Exposure to Study Medication

Exposure to study drug will be summarized including number of patients with one or two doses for TCZ and 1 to 10 doses for RDV. Patients who are dosed with RDV prior to randomization will have those doses counted.

A listing of patients by treatment group will be prepared detailing dosing of study drugs, volume administered and any dose modification due to incomplete dosing or skipped doses of TCZ/PBO or RDV not due to death or discharge.

4.6.2 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.

AEs will be coded and tabulated by system organ class, and/or preferred term (PT) and treatment arm. In tabulations, PTs and their associated system organ class will be presented in order of descending frequency summed across the treatment arms. Summaries will be produced up to Day 28 and up to Day 60.

AEs 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:

- SAEs
- AEs leading to withdrawal of study drug
- AEs leading to discontinuation from the study
- AEs leading to death

Adverse events of special interest will be defined using system organ class, 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 system organ class)
- Opportunistic infections (Roche Standard AEGT Basket)
- 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 or remdesivir infusion; and a separate summary using the Anaphylactic Reaction SMQ Narrow for events occurring during or within 24 hours of the end of tocilizumab or remdesivir infusion)

- Gastrointestinal perforations (Gastrointestinal perforation SMQ Wide)
- Bleeding events (Hemorrhages SMQ Wide)
- Demyelinating events (Demyelination SMQ Narrow)A glossary showing the mapping of investigator verbatim terms to PTs will be produced for all AEs included in the analysis. For each AESI table based on SMQs/AEGTs, a corresponding listing of the PTs that comprise the SMQ will be produced.

Listings of AEs and SAEs will be produced. AESIs will also be listed.

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

4.6.3 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. Summaries of the number of patients by CTC grade for hematology (WBC, hemoglobin, platelets), liver function tests (alkaline phosphatase, ALT, AST, total bilirubin), and renal function (creatinine, BUN) will be produced. The number of patients will be summarized by CTCAE grade category for baseline and worst postbaseline result.

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

4.6.4 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 as defined in Section 4.4. Additionally, a graphical representation of means over time of oxygen saturation and temperature (daily to Day 28) 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 respiratory support will be summarized over time, including type of support given by treatment group. Non-invasive positive pressure ventilation and mechanical ventilation will be summarized by treatment group.

4.7 INTERIM ANALYSES

No interim analyses for efficacy have been conducted during the course of this trial. See Section 3.2 for description of DMC safety monitoring.

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

Calkins KL, Canan CE, Moore RD, et al. An application of restricted mean survival time in a competing risks setting: comparing time to ART initiation by injection drug use. BMC Med Res Methodol 2018;18(27). https://doi.org/10.1186/s12874-018-0484-z.

Dmitrienko A and Tamhane A. Chapter 5. Gatekeeping Procedures in Clinical Trials. In Multiple Testing Problems in Pharmaceutical Statistics. Dmietrienko et al Editors. 2010 Routledge.

Fine P and Robert G. A Proportional Hazards Model for the Subdistribution of a Competing Risk, J Am Stat Assoc 1999;94:446, 496-509, DOI: 10.1080/01621459.1999.10474144

Guaraldi G, Meschiari M, Cozzi-Lepri A, et al. Tocilizumab in patients with severe COVID-19: a retrospective cohort study. The Lancet Rheumatology 2020;2(8):e474-84.

Gupta S, Wang W, Hayek SS, et al. Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19. JAMA Intern Med. Published online October 20, 2020. doi:10.1001/jamainternmed.2020.6252

Hermine O, Mariette X, Tharaux PL, et al. Effect of tocilizumab vs usual care in adults hospitalized with COVID-19 and moderate or severe pneumonia: a randomized clinical trial. JAMA Intern Med. Published online October 20, 2020. doi:10.1001/jamainternmed.2020.6820

Mantel N. Chi-Square Tests with One Degree of Freedom, Extensions of the Mantel–Haenszel Procedure". J Am Stat Assoc.1963: 58 (303): 690–700

McCaw ZR, Tian L, Vassy JL, et al. How to Quantify and Interpret Treatment Effects in Comparative Clinical Studies of COVID-19. Ann Intern Med 2020;173(8):632-637. doi:10.7326/M20-4044

National Heart, Lung, and Blood Institute ARDS Clinical Trials Network, Truwit JD, Bernard GR, et al. Rosuvastatin for sepsis-associated acute respiratory distress syndrome. N Engl J Med 2014;370(23):2191-2200.

Perrone F, Piccirillo MC, Ascierto PA, et al. F. Tocilizumab for patients with COVID-19 pneumonia. The single-arm TOCIVID-19 prospective trial. J. Transl. Med. 2020;18(1):1-1.

Salvarani C, Dolci G, Massari M, et al. Effect of tocilizumab vs standard care on clinical worsening in patients hospitalized with COVID-19 pneumonia: a randomized clinical trial. JAMA Intern Med. Published online 2020. doi:10.1001/jamainternmed.2020.6615

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.

Stone JH, Frigault MJ, Serling-Boyd NJ, et al. Efficacy of Tocilizumab in Patients Hospitalized with Covid-19. N Engl J Med 2020.

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]. Available from: http://www.chinaxiv.org/abs/202003.00026

Appendix 1 Protocol Synopsis

TITLE:	**A PHASE III, RANDOMIZED, DOUBLE-BLIND, MULTICENTER STUDY TO EVALUATE THE EFFICACY AND SAFETY OF REMDESIVIR PLUS TOCILIZUMAB COMPARED WITH REMDESIVIR PLUS PLACEBO IN HOSPITALIZED PATIENTS WITH SEVERE COVID-19 PNEUMONIA**
PROTOCOL NUMBER:	WA42511
VERSION NUMBER:	5
EUDRACT NUMBER:	2020-002275-34
IND NUMBER:	148225
NCT NUMBER:	NCT04409262
TEST PRODUCTS:	Tocilizumab (RO4877533) Remdesivir
PHASE:	Phase III
INDICATION:	Severe COVID-19 pneumonia
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives and Endpoints

This study will evaluate the efficacy and safety of combination therapy with remdesivir plus tocilizumab compared with remdesivir plus placebo in hospitalized patients with severe coronavirus disease 2019 (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 the remdesivir plus tocilizumab arm compared with the remdesivir plus placebo arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoint:

- Time from administration of tocilizumab/placebo to hospital discharge or "ready for discharge"

  Hospital discharge or "ready for discharge" is defined as a score of 1 on the 7-point ordinal scale. Death is defined as a competing risk.

Secondary Efficacy Objective

The secondary efficacy objective for this study is to evaluate the efficacy of the remdesivir plus tocilizumab arm compared with the remdesivir plus placebo arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoints:

- Time to mechanical ventilation or death, defined as the time from administration of tocilizumab/placebo to the first occurrence of mechanical ventilation or death (whichever occurs first)
- Time to improvement, defined as the time from administration of tocilizumab/placebo to an improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical status

- Clinical status as assessed by the investigator using a 7-category ordinal scale of clinical status on Days 7, 14, 21, 28, and 60
- Proportion of patients requiring initiation of mechanical ventilation postbaseline up to Day 28 and Day 60 (patients who do not require mechanical ventilation at baseline)
- Proportion of patients who are alive and free of respiratory failure up to Day 28 and Day 60 (patients requiring mechanical ventilation at baseline)
- Duration of mechanical ventilation (patients who require mechanical ventilation at baseline)
- Time to death up to Day 28 and Day 60
- Mortality rate on Days 14, 28, and 60
- Time to recovery, defined as time from administration of tocilizumab/placebo to the time when a category of 2, non-ICU hospital ward (or "ready for hospital ward") not requiring supplemental oxygen, or better is observed

A predefined meta-analysis will be performed on all primary and secondary endpoints.

Exploratory Efficacy Objective

The exploratory efficacy objective for this study is to evaluate the efficacy of the remdesivir plus tocilizumab arm compared with the remdesivir plus placebo arm for the treatment of severe COVID-19 pneumonia on the basis of the following endpoints:

- Time to clinical improvement (TTCI), defined as time from administration of tocilizumab/placebo to National Early Warning Score 2 (NEWS2) score of ≤ 2 maintained for 24 hours:
- Time to clinical failure, defined as the time from administration of tocilizumab/placebo to the first occurrence of mechanical ventilation, ICU admission, death, or withdrawal (whichever occurs first)

  For patients entering the study already on mechanical ventilation, clinical failure is defined as a one-category worsening on the ordinal scale, withdrawal prior to discharge, or death.
- Proportion of patients requiring initiation of ICU care postbaseline up to Day 28 and Day 60
- Duration of ICU stay (days)
- Duration of supplemental oxygen use (days)
- Proportion of patients requiring initiation of vasopressor use postbaseline
- Duration of vasopressor use
- Proportion of patients requiring initiation of extracorporeal membrane oxygenation (ECMO) postbaseline
- Duration of ECMO
- Organ failure-free days from administration of tocilizumab/placebo to Day 28
- Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral load up to Day 28 and Day 60
- Time to reverse-transcriptase polymerase chain reaction (RT-PCR) SARS-CoV-2 virus negativity

Safety Objective

The safety objective for this study is to evaluate the safety of the remdesivir plus tocilizumab arm compared with the remdesivir plus placebo arm 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, Version 5.0 (NCI CTCAE v5.0)

- Proportion of patients with any post-treatment infection at specified timepoints
- Change from baseline in targeted clinical laboratory test results

Pharmacokinetic Objective

The pharmacokinetic (PK) objective for this study is to characterize the PK profile of remdesivir and metabolite(s) in patients with severe COVID-19 pneumonia on the basis of the following endpoint:

- Plasma concentration of remdesivir and metabolite(s) 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 remdesivir plus tocilizumab combination treatment (i.e., predictive biomarkers), may serve as early surrogates of efficacy or be associated with 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 provide further evidence of tocilizumab pharmacological activity (i.e., pharmacodynamic biomarkers), or could overall increase our knowledge and understanding of disease pathogenesis and drug safety, on the basis of the following endpoint:

- Relationship between biomarkers in serum, peripheral blood mononuclear cells (PBMCs), blood, and tissue and efficacy or safety endpoints

STUDY DESIGN

Description of the Study

This is a Phase III, randomized, double-blind, multicenter study to assess the efficacy and safety of remdesivir plus tocilizumab compared with remdesivir plus placebo in hospitalized patients with severe COVID-19 pneumonia. The Sponsor intends to enroll patients who have been diagnosed with severe COVID-19 pneumonia and meet the entry criteria in centers globally. Enrollment may be up to 800 patients but will target approximately 650 patients.

In addition to the randomized patients from this study, patients from the Phase III studies WA42380 and ML42528 will be included in a meta-analysis of the primary and secondary endpoints. This analysis will assist in the interpretation of the effect of tocilizumab in combination with remdesivir compared with tocilizumab alone.

Patients must be at least 12 years of age and hospitalized with confirmed 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 require supplemental oxygen > 6 L/min to maintain $SpO_2 > 93\%$.

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 an initial 2:1 ratio to receive blinded treatment of either remdesivir plus tocilizumab or remdesivir plus placebo. Study treatment will be given in combination with standard supportive care. The randomization will be stratified by geographic region (North America, Europe, Other) and a 2-level factor based on the assessment of the 7-category ordinal scale of clinical status at screening, with levels 4–5 and 6. The proportion of randomized patients in the scale = 6 stratum will also be no more than 25%.

Patients assigned to the remdesivir plus tocilizumab (RDV + TCZ) arm will receive remdesivir as a 200 mg IV loading dose followed by one infusion of tocilizumab 8 mg/kg (maximum dose of 800 mg) on Day 1. Patients will subsequently be administered a 100 mg once-daily IV maintenance dose of remdesivir from Days 2–10. Remdesivir will be discontinued at the time of hospital discharge even if 10 days of remdesivir dosing have not been completed.

Remdesivir dosing must be adjusted for patients that enter the study having received prior remdesivir. Patients who received remdesivir prior to randomization must not exceed 10 days

of dosing in total (including remdesivir received prior to the study and during the study). If a patient received 1 dose of remdesivir (200-mg loading dose) prior to Day 1, the patient should receive remdesivir 100 mg on Days 1–9. If a patient received 2 doses of remdesivir (200-mg loading dose followed by 100-mg maintenance dose) prior to Day 1, the patient should receive remdesivir 100 mg on Days 1–8. Patients who received remdesivir prior to trial entry must not receive a second 200-mg loading dose on Day 1. The same sequence and timing of remdesivir and tocilizumab will be followed on Days 1 and 2 regardless of whether the patient has received remdesivir prior to randomization.

For both arms, if the patient has a sustained fever or clinically significant worsening of signs or symptoms (e.g., an increased supplemental oxygen requirement), one additional infusion of blinded tocilizumab/placebo can be given 8–24 hours after the first tocilizumab/placebo infusion. The second dose of blinded tocilizumab must not be given if the patient develops an adverse event or laboratory abnormalities that warrant discontinuation of tocilizumab.

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 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, chest X-ray, ECG, 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 drug.

If patients are discharged from hospital prior to Day 28, follow-up visits should be conducted on Day 14, Day 21, and Day 28 (± 2 days). Follow-up visits on Day 14 and Day 21 may be conducted as telephone visits. Patients should return to the site for the Day 28- visit, if at all possible. 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 for discharged patients. Patients should return to the site for the Day 60 visit, if at all possible.

Number of Patients

The target enrollment for the study is approximately 650 hospitalized patients with severe COVID-19 pneumonia, but up to 800 patients may be enrolled to achieve approximately 520 patients discharged by Day 28.

Target Population

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Documented informed consent to participate in the study
- Signed Assent Form when appropriate, as determined by the patient's age and individual site and country standards
- Age ≥ 12 years at time of informed consent
- 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

  The PCR test must be collected ≤ 7 days before randomization.
- Requiring > 6 L/min supplemental oxygen to maintain SpO_2 > 93%
- Agrees to not participate in another clinical trial for the treatment of COVID-19 while participating in this study
- 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.

Hormonal contraceptive methods must be supplemented by a barrier method.

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 tocilizumab or other monoclonal antibodies
- Known hypersensitivity to remdesivir, the metabolites, or formulation excipients
- 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
- Treatment with immunosuppressive or immunomodulatory therapy (including tocilizumab) within the past 3 months
- Concurrent treatment with other agents with actual or possible direct-acting antiviral activity against SARS-CoV-2 within 24 hours prior to study drug dosing

  In addition, patients with prior or current treatment with > 2 doses remdesivir for COVID-19 are excluded.
- Participating in another drug clinical trial
- Estimated glomerular filtration rate (eGFR) < 30 mL/min (including patients receiving hemodialysis or hemofiltration), using the equation described in the FDA EUA Fact Sheet for remdesivir
- ALT or AST > 5 × upper limit of normal (ULN) detected within 24 hours of screening (according to local laboratory reference ranges)
- ANC < 1000/μL at screening
- Platelet count < 50,000/μL at screening

- Body weight < 40 kg
- Pregnant or breastfeeding, or positive pregnancy test in a predose examination
- Treatment with an investigational drug within 5 half-lives or 30 days (whichever is longer) of randomization
- 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 occurs or the date at which the last data point required for the 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 8 months.

Investigational Medicinal Products

Test Products (Investigational Drugs)

The investigational medicinal products (IMPs) for this study are remdesivir IV, tocilizumab IV, and placebo for tocilizumab IV

Patients assigned to the remdesivir plus tocilizumab (RDV + TCZ) arm will receive remdesivir loading dose followed by one infusion of tocilizumab on Day 1, and once-daily maintenance dose of remdesivir from Days 2–10.

Patients assigned to the remdesivir plus placebo (RDV + placebo) arm will receive remdesivir as loading dose followed by one infusion of tocilizumab-placebo on Day 1, and once-daily maintenance dose of remdesivir from Days 2–10.

For both arms, if the patient has a sustained fever or clinically significant worsening of signs or symptoms (e.g., an increased supplemental oxygen requirement), one additional infusion of blinded tocilizumab/placebo can be given 8–24 hours after the first tocilizumab/placebo infusion. The second dose of blinded tocilizumab must not be given if the patient develops an adverse event or laboratory abnormalities that warrant discontinuation of tocilizumab.

Administration of study treatments will be performed in a monitored setting where there is immediate access to trained personnel and adequate equipment and medicine to manage potentially serious reactions.

Statistical Methods

Primary Analysis

The primary efficacy objective for this study is to evaluate the efficacy of the remdesivir plus tocilizumab arm compared with remdesivir plus placebo arm using the following endpoint:

- Time from administration of tocilizumab/placebo to hospital discharge or "ready for discharge"

  Hospital discharge or "ready for discharge" is defined as a score of 1 on the 7-point ordinal scale. Death is defined as a competing risk.

The distribution of time from administration of tocilizumab/placebo to hospital discharge (or ready for discharge) will be compared with the remdesivir plus tocilizumab arm and the remdesivir plus placebo arm up to Day 28. Patients discharged after Day 28 will be administratively censored. The distributions will be compared using an appropriate method for comparing censored event distributions with a competing risk such as the Cox model with cause specific hazards. The competing risk of mortality will be assessed over the 28-day period and addressed in the analysis of key secondary endpoints and sensitivity analysis (for details, see the SAP). Cumulative incidence plots will be presented as well as median time

to discharge (or ready for discharge), with 95% confidence intervals for the remdesivir plus tocilizumab arm and the remdesivir plus placebo arm.

Determination of Sample Size

The primary endpoint, time to discharge or "ready for discharge," is event driven. Based on the severe cohort receiving 10 days of remdesivir in Gilead's SIMPLE trial (Study GS-US-540-5773), the median time to discharge or "ready for discharge" was 11 days. Assuming a median time to discharge or "ready for discharge" of 11 days in the remdesivir plus placebo arm, a hazard ratio of 1.3 or an approximately 2.5-day reduction in median time for remdesivir plus tocilizumab versus remdesivir plus placebo, and a 2:1 randomization to remdesivir plus tocilizumab or remdesivir plus placebo, approximately 650 patients are needed to accrue approximately 520 events to achieve approximately 80% power. A reduction of at least 2 days in median time to discharge was considered clinically meaningful and the sample size was increased to ensure that a minimum effect size could be detected. Further sample size adjustments may be considered during the study based on external information. The sample size may be increased up to a maximum of approximately 800 randomized patients if fewer events than expected are observed or further shifts in standard of care warrant reassessing sample size assumptions.

Interim Analyses

There will be up to three optional interim analyses. The first interim analysis can occur after approximately one-third to one-half of the patients have been assessed for the primary endpoint on Day 28, depending on enrollment rate. There may be up to two additional unplanned interim analyses, and these will be considered if there are major changes to the study design following the first interim analysis.

Full statistical details of any optional interim analyses, along with the rationale and timing will be documented in an interim SAP, which will be made available to the relevant health authorities before the data snapshot for the first interim analysis. Only the Data Monitoring Committee (DMC) will be unblinded and decision criteria will be specified in the interim SAP. Should an unblinded interim analysis occur, any necessary adjustment to the type 1 error rate will be specified in the SAP prior to unblinding the study.

The optional interim analysis will be based on the time to discharge (or ready for discharge) and mortality. Questions to be addressed at the interim analysis might include futility as well as potential efficacy.

The DMC will also evaluate safety according to policies and procedures detailed in the DMC Charter. Regular safety reviews will begin after approximately 45 patients (15 remdesivir plus placebo and 30 remdesivir plus tocilizumab) have been enrolled and have reached 14 days of 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 DMC initially consisted of Sponsor representatives not involved in any operational aspects of the study and a Scientific Oversight Committee of external experts (responsibilities and operating principles of the DMC are described in a charter, the Internal Monitoring Committee and Scientific Oversight Committee Agreement). The DMC responsibilities transitioned to a fully independent Data Monitoring Committee (iDMC) prior to the third scheduled safety review after approximately 300 patients reached Day 28 of follow-up. The iDMC will also conduct a fourth scheduled safety data review when at least 450 patients have reached Day 28. Data processing will be handled by a Sponsor statistician and statistical programmer independent of the study management team. Details are reported in the iDMC Charter.

Appendix 2 Schedule of Activities: Screening, Days 1 and 2

Study Day	Screening [a, b]	Day 1			Day 2	
	−2 to 0					
Time After Initial Treatment (Assessment Window)		Baseline 0 Predose	1 hour After End of RDV Infusion (+ 2 hrs)	15 min After End of TCZ Infusion (+ 3 hrs)	24 hrs (± 4 hrs)	36 hrs (± 4 hrs)
Informed consent	x					
Inclusion/exclusion criteria	x	x				
Demographic data	x					
Randomization		x				
Medical history	x					
Complete physical examination [c]	x					
Weight [d]	x					
COVID-19 diagnosis [e]	x					
Chest X-ray/CT scan [f]	x					
ECG	x					
Pregnancy test [g]	x					
PaO_2/FiO_2 [h]	Optional [h]	← Optional →				
SpO_2 [i]	x	x	x	x	x	x
Vital signs [i]	x	x	x	x	x	x
Ordinal scoring [j]		x			x	
Adverse events [k]	x [k]	x [k]	x	x	x	x
Concomitant medications [l]	x	x			x	

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

Study Day	Screening [a, b]	Day 1			Day 2	
	–2 to 0					
Time After Initial Treatment (Assessment Window)		Baseline 0 Predose	1 hour After End of RDV Infusion (+ 2 hrs)	15 min After End of TCZ Infusion (+ 3 hrs)	24 hrs (± 4 hrs)	36 hrs (± 4 hrs)
Hematology [m]	x	x			x	
Chemistry (full panel) [n]	x	x				
Chemistry (only ALT, AST, total bilirubin, ALP, and creatinine)					x	
Coagulation [o]	x	x				
RDV administration [p]		x			x	
TCZ/placebo administration [q]			x			
Central Labs						
Serum sample for exploratory biomarkers [p]		x			x	
Nasopharyngeal swab: SARS-CoV-2 viral load and exploratory biomarkers [r]		x			x [r]	
Serum SARS-CoV-2 antibody titer		x				
Cryopreserved PBMCs [s]		x			x	
Blood in PAXgene® tubes for RNA analyses [t]		x				

CT = computed tomography; eCRF = electronic case report form; hrs = hours; min = minute; NEWS2 = National Early Warning Score 2; NK = natural killer (cell); PaO_2/FiO_2 = arterial oxygen partial pressure/fraction of inspired oxygen; PBMCs = peripheral blood mononuclear cells; PCR = polymerase chain reaction; PCT = procalcitonin; RDV = remdesivir; SARS-Cov-2 = severe acute respiratory syndrome coronavirus 2; SpO_2 = peripheral capillary oxygen saturation; TCZ = tocilizumab.

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

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

[a] Results from standard-of-care tests or examinations (including physical exam) performed prior to obtaining informed consent and within 24 hours before screening 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. The screening and baseline visit may be performed on the same day, provided that the patient meets all of the study entry criteria as outlined in Section 4.1.1 and Section 4.1.2 prior to randomization. If the screening and baseline visits occur on the same day, assessments do not need to be repeated.

[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, 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 after randomization should be reported as adverse events (see Section 5.3).

[d] If it is not feasible to weigh bed-bound patients, historical body weight may be used.

[e] COVID-19 test (SARS-CoV-2 PCR) to confirm diagnosis should be collected within 7 days prior to randomization.

[f] Screening chest X-ray or CT scans should be performed within 48 hours prior to randomization.

[g] For women of childbearing potential, including those who have had a tubal ligation will have a pregnancy test at screening. If a urine pregnancy test is positive, it must be confirmed with a serum pregnancy test. Study drug infusion must not be administered unless the serum pregnancy test result is negative.

[h] If arterial blood gases are measured.

[i] On Day 1, all vital sign measurements (i.e., respiratory rate, pulse rate, systolic and diastolic blood pressure, and body temperature), oxygen saturation (SpO_2), and NEWS2-specific assessments (i.e., consciousness and presence or absence of oxygen support) must be recorded 1) prior to administration of remdesivir, 2) after administration of remdesivir/prior to administration of tocilizumab/placebo and 3) after administration of tocilizumab/placebo. On Day 2, measurements should be recorded together twice daily approximately every 12 hours. If measured more than once during a 12 hour period on Day 2, 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. For patients requiring supplemental oxygen, the oxygen flow rate (L/min) and/or FiO_2 should be recorded.

[j] Assessment of clinical status using the ordinal scale should be recorded at baseline on Day 1 and again daily every morning (between 8:00 a.m. and 12:00 p.m.) for patients who remain hospitalized. See Section 4.5.10 for additional details.

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

[k] 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 first 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 (see Section 5.6).

[l] Medication (e.g., prescription drugs, over-the-counter drugs) 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.

[m] 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, or if the site can ship the blood sample on the day of collection to the central laboratory for testing. Samples are not required if neither of these options is available).

[n] Chemistry panel (serum or plasma) includes bicarbonate, sodium, potassium, chloride, glucose, BUN, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, ALP, ALT, AST, LDH, ferritin, and PCT.

[o] Coagulation panel includes INR, PTT, D-dimer and fibrinogen.

[p] Remdesivir should be administered after review of hematology and chemistry safety labs and collection of all central lab samples. The initial infusion of remdesivir should be given within 4 hours of randomization. *If the patient received remdesivir prior to randomization, the remdesivir dose on Day 1 will be a 100-mg maintenance dose.* If a second tocilizumab/placebo infusion is given, there must be at least 2 hours after completion of the second tocilizumab/placebo infusion before the Day 2 remdesivir infusion is administered – in this instance, the visit window on Day 2 at 24 hours may be extended to ±6 hours.

[q] The infusion of tocilizumab/placebo must not be given until at least 60 minutes after the infusion of remdesivir is complete. If the patient has a sustained fever or clinically significant worsening of signs or symptoms (e.g., an increased supplemental oxygen requirement), one additional infusion of blinded tocilizumab/placebo can be given 8–24 hours after the first tocilizumab/placebo infusion (see Section 5.1.3.1).

[r] Viral load will be assessed by nasopharyngeal swab or other respiratory sample (e.g., endotracheal tube aspirate). Sample collection on Day 2 is optional. Where possible the same nostril should be used. Patients who are intubated and undergo bronchoscopy will have samples taken for virological assessment. Genotypic and phenotypic viral resistance may be assessed in exploratory analyses. Other analyses may include evaluation of levels of inflammatory biomarkers.

[s] For sites capable of performing PBMC isolation and cryopreservation. If PBMC isolation or cryopreservation is not feasible for the site, then blood samples may alternatively be shipped on the day of collection to the central laboratory. Samples are not required if neither of these options is available.

[t] The PAXgene® blood RNA tube should be the last tube draw in the phlebotomy procedure to avoid contact with RNA preservation reagent inside the tube.

Appendix 3 Schedule of Activities: Days 3–28

	Days 3–28 [a]																									Study Discontinuation [b]
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	
Vital signs [c]	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
SpO_2 [c]	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 [d]	← Optional →																										Optional
Ordinal scoring [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
Adverse events [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
Concomitant medications [g]	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 [h]	x		x		x			x				x							x							x	x
Chemistry (full panel) [i]	x				x			x				x							x							x	x
Chemistry (only ALT, AST, total bilirubin, ALP and creatinine)		x	x	x		x	x																				
Coagulation [j]					x							x							x							x	x
RDV administration [k]	x	x	x	x	x	x	x	x																			
Central Labs																											
Plasma PK [l]		x			x																						
Serum sample for exploratory biomarkers	x				x							x							x							x	x
Nasopharyngeal swab: SARS-CoV-2 viral load and exploratory biomarkers [m]	x	x [m]	x	x [m]	x			x				x							x							x	x
Serum SARS-CoV-2 antibody titer																										x	x

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

Cryopreserved PBMCs [n]	x				x						x						x							x	x
Blood in PAXgene® tubes for RNA analyses [o]	x				x																			x	x

NEWS2 = National Early Warning Score 2; NK = natural killer (cell); PaO_2/FiO_2 = arterial oxygen partial pressure/fraction of inspired oxygen; PBMCs = peripheral blood mononuclear cells; PCT = procalcitonin; PK = pharmacokinetic; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2; SpO_2 = peripheral capillary oxygen saturation.

[a] If patients are discharged from hospital prior to Day 28, follow-up visits should be conducted on Day 14, Day 21 and Day 28 (±2 days). Follow-up visits on Day 14 and Day 21 may be conducted as telephone visits. Vital signs and laboratory testing will not be required for telephone visits. For patients discharged with supplemental oxygen prior to Day 28, SpO_2 measured by the patient at home should be recorded during telephone visits, if available. Patients should return to the site for the Day 28 visit*, if at all possible*.

[b] Patients who discontinue from study treatment should continue in the study and complete all assessments through Day 60 (as outlined in Appendix 2 and Appendix 3).

[c] All vital sign measurements (i.e., respiratory rate, pulse rate, systolic and diastolic blood pressure, 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 approximately every 12 hours 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.

[d] If arterial blood gases are measured.

[e] Assessment of clinical status using the ordinal scale should be recorded daily every morning (between 8:00 a.m. and 12:00 p.m.) for patients who remain hospitalized. See Section 4.5.10 for additional details.

[f] All adverse events will be reported until 60 days after the first 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 (see Section 5.6).

[g] Medication (e.g., prescription drugs, over-the-counter drugs, vaccines, 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.

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

[h] 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, or if the site can ship the blood sample on the day of collection to the central laboratory for testing. Samples are not required if neither of these options is available). Hematology labs will not be performed if follow-up visits are conducted by telephone.

[i] Chemistry panel (serum or plasma) includes bicarbonate, sodium, potassium, chloride, glucose, BUN, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, ALP, ALT, AST, LDH, ferritin, and PCT. Chemistry labs will not be performed if follow-up visits are conducted by telephone.

[j] Coagulation panel includes INR, PTT, D-dimer and fibrinogen. Coagulation labs will not be performed if follow-up visits are conducted by telephone.

[k] Remdesivir should be administered after review of chemistry safety labs and collection of all central lab samples. If a patient is discharged from hospital before Day 10, remdesivir will be discontinued at that time. *Patients who received remdesivir prior to randomization will not exceed 10 days of remdesivir dosing in total.*

[l] On Day 4 and Day 7, PK samples should be drawn prior to the start of remdesivir infusion (predose) and also 30–60 minutes after the start of the remdesivir infusion if possible. All PK samples must be drawn from the opposite arm as the remdesivir infusion. PK samples are not required if collection and processing is not feasible.

[m] Patients who remain in hospital will have viral load assessed by nasopharyngeal swab or other respiratory sample (e.g., endotracheal tube aspirate). Sample collection on Day 4 and Day 6 is optional. Where possible the same nostril should be used. Genotypic and phenotypic viral resistance may be assessed in exploratory analyses. Other analyses may include evaluation of levels of inflammatory biomarkers.

[n] For sites capable of performing PBMC isolation and cryopreservation. If PBMC isolation or cryopreservation is not feasible for the site, then blood samples may alternatively be shipped on the day of collection to the central laboratory. Samples are not required if neither of these options is available.

[o] The PAXgene® blood RNA tube should be the last tube draw in the phlebotomy procedure to avoid contact with RNA preservation reagent inside the tube

Appendix 4 Schedule of Activities: After Day 28

			Study Completion/ Discontinuation
Study Day (Assessment Window)	35 [a] (±3 days)	45 [a] (±3 days)	60 (±3 days)
Vital signs [b]	x	x	x
SpO_2 [b]	x	x	x
Ordinal scoring [c]	x	x	x
Adverse events [d]	x	x	x
Concomitant medications [e]	x	x	x
Hematology [f]	x	x	x
Chemistry [g]	x	x	x
Coagulation [h]	x	x	x
Central Labs			
Nasopharyngeal swab: SARS-CoV-2 viral load and exploratory biomarkers [i]	x [i]	x [i]	x
Serum sample for exploratory biomarkers	x		x
Serum SARS-CoV-2 antibody titer			x
Cryopreserved PBMCs [j]			x
Blood in PAXgene® tubes for RNA analyses [k]			x

NK = natural killer (cell); PMBCs = peripheral blood mononuclear cells; PCT = procalcitonin; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2; SpO2 = peripheral capillary oxygen saturation.

[a] Patients should have follow up visits on Day 35, Day 45, and Day 60 (±3 days). The Day 35 and Day 45 visits may be conducted by telephone for discharged patients. Vital signs and laboratory testing will not be required for telephone visits. For patients discharged with supplemental oxygen, SpO_2 measured by the patient at home should be recorded during telephone visits, if available. Patients should return to the site for the Day 60 visit, *if at all possible*.

[b] For patients who remain in hospital, vital sign measurements and NEWS2-specific assessments should be conducted twice daily. 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.

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

[c] Assessment of clinical status using the ordinal scale should be recorded daily every morning (between 8:00 a.m. and 12:00 p.m.) for patients who remain hospitalized.

[d] All adverse events will be reported until 60 days after the first 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 (see Section 5.6).

[e] Medication (e.g., prescription drugs, over-the-counter drugs) 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.

[f] 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, or if the site can ship the blood sample on the day of collection to the central laboratory for testing. Samples are not required if neither of these options is available). Hematology labs will not be performed if follow-up visits are conducted by telephone.

[g] Chemistry panel (serum or plasma) includes bicarbonate, sodium, potassium, chloride, glucose, BUN, creatinine, total protein, albumin, phosphorus, calcium, total bilirubin, ALP, ALT, AST, LDH, ferritin, and PCT. Chemistry labs will not be performed if follow-up visits are conducted by telephone.

[h] Coagulation panel includes INR, PTT, D-dimer, and fibrinogen. Coagulation labs will not be performed if follow-up visits are conducted by telephone.

[i] Patients who remain in hospital will have viral load assessed by nasopharyngeal swab or other respiratory sample (e.g., endotracheal tube aspirate). Sample collection on Day 35 and Day 45 is optional. Where possible the same nostril should be used. Genotypic and phenotypic viral resistance may be assessed in exploratory analyses. Other analyses may include evaluation of levels of inflammatory biomarkers.

[j] For sites capable of performing PBMC isolation and cryopreservation. If PBMC isolation or cryopreservation is not feasible for the site, then blood samples may alternatively be shipped on the day of collection to the central laboratory. Samples are not required if neither of these options is available.

[k] The PAXgene® blood RNA tube should be the last tube draw in the phlebotomy procedure to avoid contact with RNA preservation reagent inside the tube.

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	

SpO_2 = oxygen saturation.

The oxygen saturation should be scored according to either the SpO_2 Scale 1 or 2 presented in the table above. The SpO_2 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 SpO_2 Scale 2 should be made by the treating physician and should be recorded in the eCRF. In all other circumstances, the SpO_2 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.

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 (^oC)	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.