Master Protocol for Immune Modulators for Treating COVID-19

Protocol: ACTIV-1 IM

STATISTICAL ANALYSIS PLAN (SAP)

Protocol Version 2.0

Protocol Version Date: 02 DEC 2020

SAP Version Number: 5.0

11 APR 2022

Daniel K. Benjamin, MD, MPH, PhD

Kiser-Arena Distinguished Professor, Duke University Pediatrics

Duke Clinical Research Institute

300 W. Morgan Street, Suite 800

Durham Center

Durham, NC 27701

Tel: 919-668-7081

E-mail: danny.benjamin@duke.edu

SAP Approval Page

Study Title: Master Protocol for Immune Modulators for Treating COVID-19

Version: 5.0

Date of Issue: 11 APR 2022

Study Sponsor: Daniel K. Benjamin, MD, MPH, PhD

Kiser-Arena Distinguished Professor, Duke University Pediatrics

Duke Clinical Research Institute

300 W. Morgan Street, Suite 800 / Durham Centre / Durham, NC 27701

We, the undersigned, have read and approve of this SAP and agree on its content.

_____

Sponsor Representative – Daniel K. Benjamin, MD, MPH, PhD

_____

Sponsor Representative – Soju Chang, MD, MPH

_____

ACTIV-1 Protocol Statistician – Lisa LaVange, PhD

_____

ACTIV-1 Protocol Chair – William G. Powderly, MD

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
ADaM	Analysis Data Model
AE	Adverse Event
ALT	Alanine transaminase
AST	Aspartate transaminase
ATC	Anatomical Therapeutic Classification
BMI	Body Mass Index
CDISC	Clinical Data Interchange Standards Consortium
CI	Confidence Interval
COVID-19	Coronavirus
CRP	C-Reactive Protein
CTCAE	Common Terminology Criteria for Adverse Events
DAIDS	Division of AIDS
DCRI	Duke Clinical Research Institute
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
eCRF	Electronic Case Report Form
EOS	End of Study
ET	Early Termination
FU	Follow-up
ITT	Intent-to-Treat
LLN	Lower Limit of Normal
MedDRA	Medical Dictionary of Regulatory Activities
MITT	Modified Intent-to-Treat

NEWS	National Early Warning Score
PT	Preferred Term
Q1	25th Percentile (1st Quartile)
Q3	75th Percentile (3rd Quartile)
RR	Respiratory Rate
RRR	Recovery Rate Ratio
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SE	Standard Error
SoC	Standard of Care
ULN	Upper Limit of Normal
WBC	White Blood Cell

1.0 STUDY AND DOCUMENT OVERVIEW

ACTIV-1 IM is a master protocol designed to evaluate multiple investigational agents for the treatment of moderately or severely ill patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). ACTIV-1 IM is a randomized, controlled, double-blinded trial which will evaluate each agent with respect to speed of recovery, mortality, illness severity, and hospital resource utilization. Each agent will be evaluated as add-on therapy to the standard of care (SoC) in use at the local clinics. The SoC may change during the course of the study based on other research findings. Comparisons of the investigational agents among themselves is not a research objective.

This document describes the planned statistical analyses that will be conducted at the completion of the ACTIV-1 IM study for each of the investigational agents. Once the initial unblinded data review has occurred, only the blinded statistics team members will be allowed to modify the SAP (see **Appendix 1** for the list of blinded and unblinded statisticians).

The study population corresponds to moderately and severely ill adults (≥18 years old) with the coronavirus disease 2019 (COVID-19) virus. Recruitment will target patients already hospitalized for treatment of COVID-19 infection as well as patients being treated for COVID-19 infection in Emergency Departments while waiting to be admitted to the hospital. Patients both in and out of the intensive care unit (ICU) will be included in the study population. Enrollment began in October 2020 with 3 agents selected for initial evaluation. Up to 5 agents may be evaluated in ACTIV-1 IM.

The study period is 29 days, with assessments on each day of the hospital stay. If the participants are discharged from the hospital, they will have a study visit at Days 8, 11, 15, 22, and 29. For discharged participants, it is preferred that the Day 8, 11, 15, and 29 visits are in person to obtain safety laboratory tests and blood (serum and plasma) samples for secondary research (Day 29) as well as clinical outcome data. Clinical outcome is assessed for the previous day; for example, study visit Days 8, 11, 22, and 29 represent data for Days 7, 10, 21, and 28. However, infection control or other restrictions may limit the ability of the participant to return to the clinic. In this case, these visits may be conducted by phone, and only clinical data will be obtained. The Day 22 visit does not have laboratory tests or collection of samples and is conducted by phone. There will also be a safety and clinical status assessment at 60 days conducted by phone. Treatment periods may vary by agent.

The trial is adaptive in that interim analyses are planned to assess the futility of each agent, with the goal of discontinuing those with lower probabilities of success to more effectively utilize trial resources for the remaining agents. Additionally, interim analyses are planned for early stopping for efficacy.

Alpha spending functions are used to appropriately control the probability of making an erroneous conclusion at the interim and final analyses. Safety monitoring will be performed throughout the trial, and formal stopping rules for each agent will be adopted. The Data and Safety Monitoring Board (DSMB) established for ACTIV-1 IM will have oversight responsibility for the study.

The effectiveness of each therapeutic agent as add-on therapy to SoC will be evaluated based on the primary endpoint of time to recovery through Day 28. The sample size requirements are based on the ability to detect a moderate improvement in time to recovery for each agent. A total of 788 recoveries are required for each comparison to provide approximately 85% power to detect a recovery rate ratio (RRR) of 1.25. Assuming 73% of participants achieve recovery within 28 days, consistent with the ACTT-1 (Beigel, Tomashek et al. 2020) results, the total sample size to evaluate 1, 2, and 3 agents in ACTIV-1 IM is approximately 1080, 1620, and 2160, respectively. Because each agent is being compared to SoC plus placebo with no between-agent comparisons, no multiplicity adjustments for multiple agents are planned.

2.0 STUDY OBJECTIVES AND ESTIMANDS

Primary Objective:

- To evaluate the clinical efficacy of different investigational therapeutics relative to the control arm in adults hospitalized with COVID-19 with respect to time to recovery through Day 28 from randomization.

Key Secondary Objectives:

- To evaluate the clinical efficacy of different investigational therapeutics relative to the control arm in adults hospitalized with COVID-19 according to:
  - Clinical status (8-point ordinal scale) assessed on Day 14 and Day 28
  - Mortality

Summaries of the questions of interest and descriptions of the estimand attributes are shown in **Table 1**.

Table 1. Study Objectives and Estimands.

Question of Interest	**Objective Description / Study Population**	**Endpoint**	**Intercurrent Events**	**Population Summary**
Does drug X reduce the time-to-recovery compared with standard-of-care alone for patients hospitalized with moderate or severe COVID-19?	Primary Objective of the Study / All Randomized Participants	Time to recovery defined as the first day with an 8-point ordinal scale score of ≥ 6.	Death from any cause (a semi-competing event). Once an individual has experienced death, they no longer have the opportunity to 'recover'. All other intercurrent events will be ignored.	Recovery rate ratio (RRR) and its associated confidence interval from a Fine-Gray model. Values greater than 1.0 indicate a benefit for drug X. A value of 2.0 suggests a larger benefit than 1.25. Stratification factors will include region and baseline severity (2, 3, 4, or 5). Other covariates in the model will include age and sex.
Does drug X improve the ordinal response of the 8-point ordinal scale compared with standard-of-care alone?	Key Secondary Objective of the Study / All Randomized Participants*	The 8-point ordinal scale score on Day 14. A supplemental analysis will be completed on the same scale at Day 28.	Deaths are placed as the worst possible score on the ordinal scale. All other intercurrent events will be ignored in these analyses.	Odds ratio estimate based on a proportional odds model. Values above 1.0 suggest benefit from drug X. The model will include baseline severity (2, 3, 4, or 5), region, sex, and age (continuous) as covariates.
Does drug X reduce the rate of death over 28 days compared with standard-of-care alone?	Key Secondary Objective of the Study / All Randomized Participants*	Death observed between randomization through Day 28	All other intercurrent events will be ignored in these analyses.	An odds ratio from a logistic regression model with indicator variables for treatment group,

				8-point ordinal scale (2, 3, 4, or 5) at baseline, region, sex, and age (continuous).

*For the test agent and pooled control group

Additional Secondary Objectives:

- To evaluate the clinical efficacy of different investigational therapeutics relative to the control arm as assessed by clinical severity and hospitalization.
- To evaluate clinical status at Day 60.
- To evaluate the safety of different investigational therapeutics relative to the control arm.
- To evaluate the impact of different investigational therapeutics relative to the control arm of extrapulmonary manifestations of COVID-19.

Exploratory Objectives:

- To assess the National Early Warning Score (NEWS).

3.0 STATISTICAL HYPOTHESES AND ENDPOINTS

- The primary null hypothesis being tested is that time-to-recovery does not differ between each test agent (plus SoC) and the pooled control group, consisting of patients receiving SoC plus placebo. The alternative hypothesis is that the test agent and pooled control group differ in time-to-recovery with the hope that the test agent group has a shorter time-to-recovery than the pooled control group although a two-tailed test will be used. If Drug A, B, and C denote the initial set of agents to be tested in ACTIV-1 IM plus SoC, and S denotes patients receiving SoC plus placebo, then the primary hypotheses are:

H_{01}: $TR_A = TR_S$ vs H_{A1}: $TR_A \neq TR_S$

H_{02}: $TR_B = TR_S$ vs H_{A2}: $TR_B \neq TR_S$

H_{03}: $TR_C = TR_S$ vs H_{A3}: $TR_C \neq TR_S$

- **Primary Efficacy Endpoint -** Time to recovery through Day 28

  Day of recovery is defined as the first day on which the participant satisfies one of the following three categories from the 8-point ordinal scale (defined in Section 3.1.2):
  - Hospitalized, not requiring supplemental oxygen – no longer requires ongoing medical in-patient care;
  - Not hospitalized, limitation on activities and/or requiring home oxygen;
  - Not hospitalized, no limitations on activities.

- There are two key secondary hypotheses that will be evaluated as supportive evidence.

- **Key Secondary Endpoint 1** - Clinical status (8-point ordinal scale) on Day 14

  The null hypothesis to be tested is that the odds of improvement on the 8-point ordinal scale is the same for the placebo and experimental treatment arms (i.e., the common odds ratio is 1) on Day 14. For this comparison, the parameter of interest is the "common odds ratio," which quantifies the shift in the severity distribution resulting from treatment. The odds ratio will be defined such that a value greater than 1 implies better outcomes with active drug compared to placebo control.

- **Key Secondary Endpoint 2** - Mortality

  The null hypothesis to be tested is that there is no difference in mortality rates between treatment groups. Mortality through Day 28 will be analyzed using a logistic regression model with an indicator variable for the treatment group, 8-point ordinal scale (2, 3, 4, or 5) at baseline, region, sex, and age.

- Additional Secondary Endpoints
  - Time to an improvement of one category and two categories from Day 0 (baseline) using the 8-point ordinal scale.
  - Mean change in the 8-point ordinal scale from baseline to Days 2, 4, 7, 10, 14, 21, 28, and 60.
  - Differences in total severity score (TSS): 8-point ordinal scale summarized as a daily score (for days collected) averaged over time from baseline through Day 28.
  - Days alive and free of supplemental oxygen through 28.
  - Incidence and duration of new oxygen use during the study. This includes new supplemental oxygen, non-invasive ventilation/high flow oxygen use, and invasive mechanical ventilation/extracorporeal membrane oxygenation (ECMO).
  - Days alive and free of non-invasive ventilation/high flow oxygen use through Day 28.
  - Incidence and duration of new non-invasive ventilation or high flow oxygen use during the study. This includes new non-invasive ventilation or high flow oxygen use as well as invasive mechanical ventilation/ECMO.
  - Days alive and free of invasive mechanical ventilation/ECMO through Day 28.
  - Incidence and duration of new mechanical ventilation/ECMO use during the study.
  - Days alive and out of hospitalization through Day 28.

- Key Safety Endpoints
  - Serious Adverse Events (SAEs)
  - Grade 3 and 4 Adverse Events (AEs)
  - Baseline laboratory tests include CBC with differential (including absolute neutrophil count and absolute lymphocyte count), ALT, AST, ALP, total bilirubin, creatinine (with calculated eGFR), glucose, Troponin, PT/INR, d-dimer, and C-reactive protein. *On Days 3, 5, 8, 11, 15, and 29, safety laboratory tests include WBC with differential, hemoglobin, platelets, ALT,*

*AST, ALP, total bilirubin, creatinine (with calculated eGFR), and glucose. On Day 8 or on day of discharge (if less than 24 hours prior to Day 8), the following **additional** laboratory assessments are performed: Troponin, PT/INR, d-dimer, and C-reactive protein.*

- Incidence of individual and "any" specified extrapulmonary manifestations up to Day 29

- Exploratory Endpoints
  - NEWS assessed daily while hospitalized (Day 1 – discharge) and on Days 15 and 29

3.1 STUDY DEFINITIONS AND DERIVED VARIABLES

3.1.1 VISIT DAY

The baseline visit is the date of randomization. Post-baseline study day visits are documenting the clinical information through the completion of the prior day (e.g. Day 2 is collecting the post-randomization data Day 1 and Day 29 is collecting data that encompasses all of Day 28). However, we will present visit day as actual day that is being assessed.

3.1.2 8-POINT ORDINAL SCALE

- The 8-point ordinal scale is the basis for the primary and key secondary clinical endpoints in the study, namely, time to recovery, improvement in disease severity, and mortality.
- The scale used in this study is as follows (from worst to best):
    - 1: Death;
    - 2: Hospitalized, on invasive mechanical ventilation or ECMO;
    - 3: Hospitalized, on non-invasive ventilation or high flow oxygen devices;
    - 4: Hospitalized, requiring supplemental oxygen;
    - 5: Hospitalized, not requiring supplemental oxygen – requiring ongoing medical in-patient care (COVID-19 related or otherwise);
    - 6: Hospitalized, not requiring supplemental oxygen – no longer requires ongoing medical in-patient care;
        - This would include those kept in hospital for quarantine/infection control/social reasons, awaiting bed in rehabilitation facility or homecare, etc.
    - 7: Not hospitalized, limitation on activities and/or requiring home oxygen;
    - 8: Not hospitalized, no limitations on activities

3.1.3 MEASURES OF CLINICAL SUPPORT, LIMITATIONS, AND INFECTION CONTROL

The participant's clinical status will be captured on each study day while hospitalized through Day 29, for the previous day. Clinical data status is measured twice on Day 1, once at the time of randomization and once at the Day 2 visit for the post-randomization time period. If a participant is discharged prior to Day 15, clinical status is collected on Day 8, Day 11, Day 15, and Day 29, if the participant returns for an in-person clinic visit or by phone if an in-person visit is not possible. Clinical status will also be captured on Day 22 during a phone visit. Clinical status is largely measured by the 8-point ordinal scale and the NEWS. Unlike the NEWS, the ordinal scale can also be evaluated over the phone if the discharged participant is unable to return for visits on Day 8, 11, 15, or 29 as well as on Day 22.

Improvement will be defined as an increase of at least one point on the 8-point scale compared to the baseline value (e.g. from 4 to 5) and the time to improvement will be defined as the elapsed time (in days) from baseline to the earliest day of observed improvement.

Any subjects that are lost to follow-up or terminated early prior to an observed improvement will be censored at the day of their last observed assessment.

3.1.4 EXTRAPULMONARY MANIFESTATIONS

The presence of the extrapulmonary manifestation of disease during the course of hospitalization will be assessed and reported at discharge. These tables will be reported with stratification by the participant's baseline clinical status. Specifically, clinical organ failure is defined by development of any one or more of the following clinical events:

a. Respiratory dysfunction:

1. Respiratory failure defined as receipt of high flow nasal oxygen, noninvasive ventilation, invasive mechanical ventilation or ECMO

b. Cardiac and vascular dysfunction:

1. Myocardial infarction
2. Myocarditis or pericarditis
3. Congestive heart failure: new onset NYHA class III or IV, or worsening to class III or IV
4. Hypotension requiring institution of vasopressor therapy

c. Renal dysfunction:

1. New requirement for renal replacement therapy

d. Hepatic dysfunction:

1. Hepatic decompensation

e. Neurological dysfunction

1. Acute delirium
2. Cerebrovascular event (stroke, cerebrovascular accident [CVA])
3. Transient ischemic events (i.e., CVA symptomatology resolving <24 hrs)
4. Encephalitis, meningitis or myelitis

f. Hematological dysfunction:

1. Disseminated intravascular coagulation
2. New arterial or venous thromboembolic events, including pulmonary embolism and deep vein thrombosis
3. Major bleeding events (>2 units of blood within 24 hours, bleeding at a critical site (intracranial, intraspinal, intraocular, pericardial, intraarticular, intramuscular with compartment syndrome, or retroperitoneal), or fatal bleeding).

g. Secondary infection / superinfection:

1. Intercurrent, at least probable, documented serious disease caused by an infection other than SARS-CoV2, requiring antimicrobial administration and care.
2. A complete list of secondary infections / superinfections is provided in the study manual of procedures.

3.1.5 NATIONAL EARLY WARNING SCORE (NEWS)

Vital signs and other clinical assessments are collected for the calculation of the NEWS, and include temperature, systolic blood pressure, heart rate, respiratory rate, O_2 saturation and level of consciousness. Vital signs collected per standard of care can be used. NEWS has demonstrated an ability to discriminate subjects at risk of poor outcomes. (Smith, Prytherch et al. 2016). This score is based on 7 clinical parameters (see **Table 2**). The NEWS is being used as an exploratory efficacy measure. The components of NEWS should be evaluated daily while hospitalized. It can be performed concurrently with the Ordinal Scale. This should be evaluated at a consistent time for each study day and prior to administration of study product. The 7 parameters can be obtained from the hospital chart or electronic medical record (EMR) using the last measurement prior to the time of assessment (including parameters collected prior to the time of consent) and a numeric score is given for each parameter (e.g., a respiratory rate [RR] of 9 breaths per minute is one point, oxygen saturation of 92% is two points). This is recorded for the day obtained (i.e., on Day 3, the vital signs and other parameters from Day 3 will be

used to obtain NEWS for Day 3). ECMO and mechanically ventilated participants should be assigned a score of 3 for RR (RR ≤8 breaths per minute) regardless of the ventilator setting. Participants on ECMO should get a score of 3 for heart rate since they are on cardiopulmonary bypass.

Table 2. National Early Warning Score (NEWS)

PHYSIOLOGICAL PARAMETERS	3	2	1	0	1	2	3
Respiration Rate (breaths per minute)	≤8 [a]		9 – 11	12 – 20		21 – 24	≥25
Oxygen Saturation (%)	≤91	92 – 93	94 – 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature (°C)	≤35.0		35.1 – 36.0	36.1 – 38.0	38.1 – 39.0	≥39.1	
Systolic Blood Pressure (mm Hg)	≤90	91 – 100	101 – 110	111 – 219			≥220
Heart Rate (beats per minute)	≤40 [b]		41 – 50	51 – 90	91 – 110	111 – 130	≥131
Level of Consciousness				A			V, P, U

Level of consciousness = alert (A), and non-alert and arousable only to voice (V) or pain (P), and unresponsive (U).

[a] *If the participant is on ECMO or invasive mechanical ventilation, they will be given a score of 3 (≤8 breaths per minute) for respiratory rate regardless of ventilator setting.*

[b] *Participants on ECMO will also receive a score of 3 (≤40 beats per minute) for heart rate.*

3.2 POWER AND SAMPLE SIZE

ACTIV-1 is an event-driven trial based on the number of observed recovery events. The effectiveness of each therapeutic agent as add-on therapy to SoC will be evaluated based on the primary endpoint of time to recovery through Day 28. The Fine-Gray approach will be used to compare each test agent + SoC vs. SoC-alone with respect to the cumulative incidence of recovery, accounting for the semi-competing risk of mortality (Fine and Gray 1999). The approach is similar to using a log-rank test on time to recovery, retaining in the risk set people who die. The two key determinants of power are the total number of events (i.e., recoveries) and the treatment-to-control recovery rate ratio (RRR).

Overall study sample size requirements are based on the number of agents being evaluated and the ability to pool control patients for analysis. Initial sample size estimates are derived assuming three agents are ready for testing at study start and will remain in the study for evaluation at the final analysis stage. If newly emergent therapies are entered into the master protocol after study start, sample size requirements will be adjusted accordingly.

Table 3 provides the numbers of recoveries and of patients required to provide 85% power for a single pairwise comparison of test drug versus control assuming a 73% recovery rate and various RRRs. The assumed 73% recovery rate is based on the Kaplan-Meier recovery rate from ACTT-1. Note that the recovery rate is the analogue of the hazard for each test agent or control treatment, and the RRR is the analogue of the hazard ratio for a test agent relative to control in this setting. Approximately 347 recoveries are required to detect a 40% increase in the rate of recovery ($\theta = 1.40$) from control. An RRR of 1.40 is similar to, but slightly higher than the figure of 1.31 reported in Cao et al. (2020) for a lopinavir/ritonavir trial that used time to improvement by 2 categories as primary endpoint (Cao, Wang et al. 2020). A total of 436 recoveries is needed for an RRR of 1.35 with 85% power.

Table 3. Number of recoveries and number of patients needed for 85% power assuming a type I error rate of 5% for various recovery ratios

Recovery rate ratio (θ)	Number of recoveries needed for 85% power	Sample Size assuming 73% recovery rate (1 test agent vs SoC)
1.25	**788**	**1,080**
1.30	570	781
1.35	436	598
1.40	347	476

The sample size requirements are based on the ability to detect a moderate improvement in time to recovery for each agent. After accounting for the proposed interim efficacy and futility monitoring rules,

a total of 788 recoveries are required for each comparison to provide approximately 85% power to detect a RRR of 1.25. Assuming 73% of participants achieve recovery in 28 days, consistent with the ACTT-1 results (Beigel, Tomashek et al. 2020), the total sample size to evaluate 1, 2, and 3 agents in ACTIV-1 IM is approximately 1080, 1620, and 2160, respectively.

The proposed power and sample size requirements also provide adequate power to detect differences in the key secondary endpoints of clinical improvement and mortality. Based on the formulas from Whitehead (1993), a sample size of 1,080 (540 participants / group)) gives approximately 95% or 85% power to detect a common odds ratio of 1.50 of clinical improvement (the observed odds ratio in the ACTT-1 trial) or 1.40 using a 2-tailed test at level α=0.05. The 28-day mortality probability in the remdesivir arm of ACTT-1 was approximately 12%. The power for comparing 28-day mortality is 90% for a 50% relative reduction (from 12% to 6%) and 80% to detect a difference of 12.0% vs. 6.8%.

3.3 RANDOMIZATION SCHEME

Randomization will be stratified by:

- Geographic Region
    - USA – Northeast
        - Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
    - USA - Midwest
        - Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, North Dakota, Nebraska, Ohio, South Dakota, Wisconsin
    - USA – South
        - Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia
    - USA – West
        - Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, New Mexico, Nevada, Oregon, Utah, Wyoming, Washington
    - Argentina
    - Brazil
    - Peru
    - Mexico
    - Additional countries will be treated as separate regions
- Severity of illness at enrollment (by ordinal scale)
    - Severe disease:

        - Hospitalized, on invasive mechanical ventilation or ECMO, or
        - Hospitalized, on non-invasive ventilation or high flow oxygen devices.
    - Moderate disease:
        - Hospitalized, requiring supplemental oxygen, or
        - Hospitalized, not requiring supplemental oxygen.

Conceptually, randomization will proceed in two steps. At the first step, each participant will be assigned (open-label) with equal probability to one of the agents available at the time the patient is enrolled and for which the patient is eligible to receive, after applying any agent-specific safety exclusions. At the second step, each participant will be assigned to either the test agent or its matching placebo in an n:1 ratio, where n = the number of agents currently active in the master protocol and for which the patient is eligible to receive.

With SoC and three agents active simultaneously, if the participant meets each of the agent-specific criteria, this procedure results in the randomization ratio of 1:1:1:1. If 3 agents are available, but a patient is only eligible to receive 2 of them, the allocation ratio at the first step will be 1:1, and at the second step will be 2:1 (agent vs placebo). Inclusion of a matching placebo for each agent enables masking of study participants and clinical personnel to treatment assignment at the second stage. Data from patients receiving SoC plus placebo will be pooled across agents for comparative analyses and hypothesis testing. Comparative analyses of a particular agent will include the subset of pooled placebo patients who enrolled concurrently with the particular agent and would have been eligible to receive the agent. That is, patients enrolled to control arm prior to a new agent entering the trial will not be included in comparative analyses of that newly added investigation agent.

3.4 INTERIM ANALYSES AND DATA MONITORING

A complete description of the interim analyses plans including mock tables, listings, and figures is included in a separate DSMB guidance document. Briefly, unblinded interim analyses are planned to (i) assess the futility of each agent, with the goal of discontinuing those with lower probabilities of success to more effectively utilize trial resources for the remaining agents and (ii) review comparative analyses for each test agent to assess early stopping for efficacy.

Alpha spending functions will be used to appropriately control the probability of making an erroneous conclusion across interim and final efficacy analyses at $\alpha = 0.05$ (two-sided) for each agent. The familywise error rates will be controlled at the two-sided 0.05 level for the primary endpoint and the two key secondary endpoints. Briefly, a gatekeeper approach will be used to test the primary hypothesis. As described in Hung, Wang and O'Neill (2007), the null hypotheses for the two key secondary endpoints are tested only if the null hypothesis is rejected for the primary endpoint. At the second level, the two key secondary endpoints are evaluated using a Pocock-type boundary with a Holm

procedure used to control the type I error for the two secondary endpoints (Tamhane, Mehta, Liu 2010; Holm 1979).

3.4.1 INTERIM SAFETY ANALYSES

Safety analyses will evaluate Grade 3 and 4 AE and SAEs by treatment arm. Safety monitoring will be ongoing. The unblinded statistical team will prepare these reports for review by the DSMB. More details will be provided in a separate DSMB guidance document.

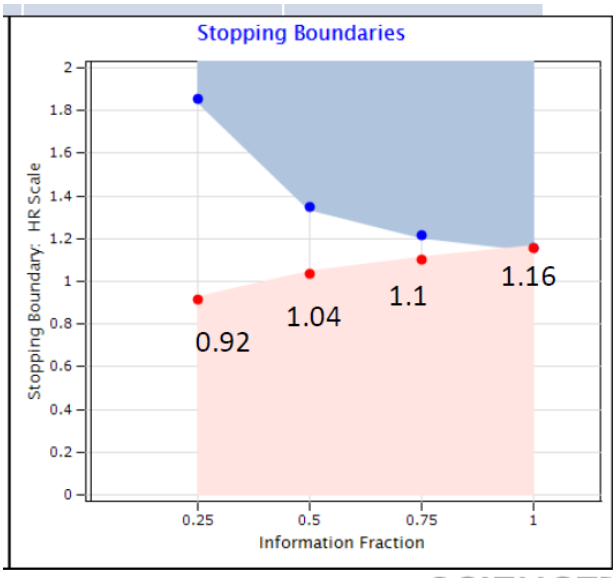
3.4.2 INTERIM ANALYSES FOR FUTILITY AND EFFICACY REVIEW

Interim analyses for futility and efficacy were planned at three times during the study corresponding to the availability of approximately 25%, 50%, and 75% of total information. The planned randomization algorithm ensures approximately equal allocation to each test agent or control assuming that participants are eligible for all available agents. It is anticipated that the interim analyses for each agent will occur at the same time, and the DSMB will make recommendations for all agents at their scheduled meetings. If recovery rates vary substantially by agent; however, it may be necessary to let the interim analysis times also vary by agent. Because Type I error probabilities are controlled for each agent, independently of the other agents, the need for additional DSMB reviews due to differential information accrual across agents should not pose any issues other than logistical ones.

The Lan-DeMets spending function analog with the O'Brien-Fleming boundaries will be used to monitor the primary endpoint as a guide for the DSMB. This spending function is conservative in that priority is given to preserving power for the final analysis with the use of stringent stopping rules early in the study.

In contrast, moderately aggressive stopping rules for futility will be implemented to promote early discontinuation of agents with low probabilities for success. The futility stopping rules will be considered non-binding by the DSMB in their review of interim data. The futility boundaries are computed based on the gamma family of spending functions (Hwang, Shih, DeCani 1990). **Figure 1** below illustrates the efficacy boundaries based on the Lan-DeMets spending function analog to O'Brien-Fleming and the futility boundaries based on a (Gamma=-2) spending function.

Figure 1: Stopping Boundaries for Efficacy (blue shading) and Futility (pink shading).



As can be seen from the figure, an RRR showing no more than a small amount of improvement in recovery time (RRR ≤ 1.04) will result in early discontinuation of the agent mid-way through information accrual in the study (**Table 4**). When the null hypothesis is true (RRR = 1), the probability of discontinuing an agent at this point is 66%, but under the alternative, this probability is 4%.

Table 4. Efficacy and futility boundaries expressed in terms of z-scores and recovery rate ratios (RRRs). Conditional power corresponding to futility boundaries is also shown, and is computed under the originally hypothesized RRR of 1.25. For the interim analyses, conditional power will be calculated under additional scenarios including the current trend and the null hypothesis.

	Efficacy Boundary		Futility Boundary		
Information Fraction	Z-score	RRR	Z-score	RRR	Conditional Power
0.25	4.3326	1.8540	-0.5854	0.9200	0.52
0.50	2.9631	1.3479	0.3791	1.0389	0.40
0.75	2.3590	1.2142	1.2187	1.1055	0.36
1.00	2.0141	1.1543	2.0141	1.1543	NA

If the primary endpoint is declared to be statistically significant in the favorable direction, the two key secondary endpoints will be tested according to a Pocock-like boundary with the Holm procedure used to control for the two key secondary endpoint comparisons (Holm 1979). See **Table 5** for the details of these procedures.

Table 5. Efficacy boundaries for the two key secondary endpoints assuming 3 equally spaced interim reviews and a final analysis.

Information Fraction	If the primary endpoint is declared significant, the larger absolute Z-statistic (smaller nominal p-value) is compared against … |Z| (one-sided p-value)	If the primary and other key secondary endpoint is declared statistically significant, the smaller absolute Z-statistic (larger p-value) is compared against … |Z| (one-sided p-value)
0.25	2.62 (0.0043)*	2.36 (0.0091)
0.50	2.62 (0.0043)	2.36 (0.0091)
0.75	2.62 (0.0043)	2.36 (0.0091)
1.00	2.62 (0.0043)	2.36 (0.0091)

*Exact boundaries will be computed using a spending function approach with a Pocock-like bound.

Please note that the DSMB meeting scheduled for the 0.25 information time occurred closer to 0.30 information fraction. Furthermore, the DSMB approval letter dated May 28, 2021 states the following …

"The NIAID DCR Data and Safety Monitoring Board held a teleconference on May 6, 2021. After carefully reviewing the safety data and interim analyses, the DSMB identified no safety concerns and recommended that all sub-studies in ACTIV-1 continue as planned. The DSMB made other recommendations as noted in the summary and recommended that the next DSMB meeting should take placed when ACTIV-1 has reached 60% enrollment."

Following the September 2, 2021 DSMB meeting, the recommendations stated that "the Board recommends that no further interim analysis be done". This meeting was the second DSMB review of the unblinded study data. 1340 participants had been randomized across 70 sites in the U.S. and Latin America at the time of the data cut.

3.4.3 ADAPTATIONS

The master protocol design with common data elements, procedures, etc., and adaptive platform design allows for seamless dropping or replacing of an arm by a new treatment. Only concurrent data will be used for comparisons. Thus, a new arm will be compared to controls enrolled on or after the time the new arm is added and determined to be eligible to receive the agent. It is anticipated that the DSMB will informally incorporate mortality and adverse event information into the decision making regarding stopping for efficacy and futility.

Prior to the second interim analysis, there is a plan to have a blinded sample size re-estimation that allows for sample size increases of as much as 25% but no reduction. This decision would be made by the members of the study team without access to the treatment codes (i.e. active or placebo) or the unblinded results. The formula of Whitehead (1993) was used to determine the expected power for the 8-point scale. That calculation includes a factor that depends on the marginal distribution of the 8-point scale lumped over the active and placebo arms. Additionally, the recommendation for the sample size adjustment will examine the overall recovery rate through Day 28. It is possible that the observed

recovery rates will be somewhat different from the ACTT-1 results due to changes in SoC, differences in participant mix, and random variability. Recommendations for the sample size adjustment will be based on trying to maintain ≥ 85% power for the primary endpoint and ≥ 70% power for the key secondary endpoints.

Approximately 2 weeks prior to the data cut for the second interim analysis, the blinded statistical team will conduct a review of marginal event rates and endpoint distributions. The goals of the blinded review are to assure that there is adequate power for the primary and two key secondary endpoints under various extreme scenarios. Namely, the team would like to maintain 85% power for the primary endpoint and >70% power for the two key secondary endpoints. The blinded statistics team will have access to limited data including the marginal 28-day mortality rate and the 14-day 8-point ordinal scale lumped across the three sub-studies. As noted in **Table 6** below, the power for these endpoints does depend on these marginal distributions.

Table 6. Outline of the Power Calculations for the Primary and Two Key Secondary Endpoints

Endpoint	**Assumptions under the Null Hypothesis** (H_0)	**Assumptions under the Alternative Hypothesis (**H_a)	**Notes**
Time-to-recovery	RRR = 1.0	RRR = 1.25	Each sub-study plans to observe >788 recovery events for the primary endpoint, which provides 85% power under the alternative hypothesis.
8-Point Ordinal Score at 14-days	Common Odds Ratio = 1.00	Common Odds Ratio = 1.50	The power for this statistical comparison is highly dependent on the marginal distribution of the 8-ordered categories. For example, if the placebo group distribution includes 12.5% for each category, a common odds ratio of 1.50 with sample sizes of 540/group will yield approximately 97% power with two-sided Type I error of 0.05. By comparison, a placebo distribution with 10% in category 1, 84% in category 8, and 1% each in categories 2 to 7, would yield only 63% power under the same conditions.
Mortality at 28 days (binary endpoint)	Odds Ratio = 1.00	Odds Ratio < 1.00 (e.g. 50% relative reduction)	12% (placebo) vs. 6% (active) with 540 participants per group will provide 88% power with Type 1 error (two-sided) of 0.025. By comparison, 6% (placebo) vs. 3% (active) will yield the only 55% power under the same conditions.

Tables 7 and 8 provide an example of the power calculations that will be conducted to determine whether there is adequate power to detect 50% reductions in 28-day mortality rates. In the example below there would be adequate power (i.e. >70% power) to detect large relative differences. With a much lower pooled mortality rate, the power would decrease.

Table 7. Four scenarios for the treatment mortality rates with marginal mortality rates of 10%.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
	No effective treatments	**1 effective treatment**	**2 effective treatments**	**3 effective treatments**
Placebo	10%	11.43%	13.33%	16.00%
Active Drug 1	10%	5.71%	6.67%	8.00%
Active Drug 2	10%	11.43%	6.67%	8.00%
Active Drug 3	10%	11.43%	13.33%	8.00%
Total	10.00%	10.00%	10.00%	10.00%

Table 8. Power calculations to detect 50% relative reductions for the 28-day mortality endpoint.

Placebo Group Rate	**Active Group Rate**	**Power with Type I Error of 0.05 (two-sided)**	**Power with Type I Error of 0.025 (two-sided)**
11.43%	5.71%	92%	86%
13.33%	6.67%	95%	92%
16.00%	8.00%	98%	96%

Assuming approximately 540 participants per group

A similar approach will be used with the marginal distribution on the 8-point scale. If these plausible scenarios provide greater than 70% power to detect a common odds ratio of 1.50, the sample size will not be increased.

3.5 DATA SOURCES

At each study site, data will be entered on the electronic Case Report Forms (eCRFs) stored in Medidata Rave. Prior to database lock, programmed computer edit checks will be run against the database to identify discrepancies and verify reasonableness of the data. Queries to resolve discrepancies will be generated and resolved by the sites. As needed, DCRI Statistics will be able to directly download the eCRF database from Rave using SAS on Demand as Clinical Data Interchange Standards Consortium Study Data Tabulation Model (CDISC SDTM version 3.2) datasets. CDISC Analysis Data Model (ADaM version 2.1) datasets will be created by DCRI Statistics for production of final tables, figures, and listings.

All planned reporting will be based off of CDISC datasets, but in the case of emergent safety data, some reporting may occur from the raw eCRF data. All programs written to create analysis datasets and perform analyses will be validated according to SOPs established by the DCRI Statistical Programming group.

3.6 DOCUMENTATION CONVENTION

The statistical analyses described in this SAP, as well as production of tables, listings, and figures will be performed using SAS®, version 9.4 or higher (SAS Institute, Cary, NC). Additional statistical software may be used as needed.

3.7 VERIFICATION OF RESULTS

All tables, listings, and graphs will be verified and reviewed before considered final. The verification process will ensure that the numbers are produced by a statistically valid method and that the execution of the computations is correct. Qualified statisticians or statistical programmers employed by the DCRI who have not been previously involved in the production of the original programming will perform the verification procedures. Methods of verification include independent programming, prior to issuance of the draft statistical report, of all analysis datasets/ADaM and comparison to data listings. Tables, listings, and figures will be reviewed for accuracy, consistency with this analysis plan, consistency within tables/listings/figures, and consistency with corresponding output. Once verification is complete, all documentation of the verification process will be filed as required by the Statistical Standard Operations Procedures of the DCRI.

3.8 SUBJECT DISPOSITION

The disposition of subjects (number randomized, number who received any amount of the randomly assigned treatment, number completing study drug administration, number who withdrew consent or discontinued from study drug early, and number lost to follow-up, and number who completed the trial) will be summarized by treatment group. The number of subjects screened for inclusion and a breakdown of reasons for exclusion will be summarized. The timing and reasons for early discontinuation of study drug and/or withdrawal from the study will be summarized by treatment group. For the calculation of percentages, subjects who die will not be included in the denominators for visits/assessments beyond their death. Treatment compliance (e.g. number of subjects who had their required infusions halted/slowed and the number of subjects with missed doses) will be summarized by

treatment group. A subject listing of analysis population eligibilities will be generated. A listing of all subjects discontinued from the study after enrollment, broken down by site and treatment group will be provided. The listing will include: reason for discontinuation, treatment group, duration of treatment, and whether or not the blind was broken. Also, for subjects who discontinued from the study after enrollment, a listing of adverse events will be provided. Subject-specific protocol deviations will be summarized by the reason for the deviation, the deviation category, treatment group, disease severity and (separately) site for all subjects. All subject-specific protocol deviations and non-subject specific protocol deviations will be included in listings.

3.9 POPULATIONS FOR ANALYSES

The primary analysis will be based on an intent-to-treat (ITT) population, including all participants randomized. Similarly, safety analyses will be based on a modified intention-to-treat (MITT) population consisting of all participants who received at least one dose or injection of each drug administered in the randomization arm (e.g. investigational agent/placebo plus SoC). Each substudy will define separate ITT and MITT populations due to the shared SoC plus placebo (or control) group. The primary analysis will be based on those participants randomized in order to achieve 788 recoveries for each pairwise comparison. Subsequent analysis will be performed on all randomized participants.

4.0 STATISTICAL ANALYSES

4.1 GENERAL APPROACH

- Statistical significance: Statistical comparisons will be performed using two-sided significance tests. An alpha level of 0.05 will determine significance (unless otherwise noted) for each agent. Analyses will be carried out separately for each test agent.

- Descriptive statistics:
    - Continuous variables will be presented as n, mean, standard deviation, median, Q1, Q3, and minimum and maximum. For comparisons of treatment groups, we will use the non-parametric Wilcoxon-Mann-Whitney rank-sum test.
    - Categorical variables will be presented as percentage (number). Group comparisons will use the conventional chi-square test.
    - Binary variables will be presented as percentage (number). Group comparisons will use the conventional chi-square test or Fisher's Exact Test.

- When death is not a competing risk (for example, the endpoint includes death in the composite), differences in time-to-event endpoints by treatment will be summarized with Kaplan-Meier curves, log-rank tests, stratified log-rank tests, and hazard ratios from Cox proportional hazards models. When death is a competing risk (for example, time to at least a one-category improvement in ordinal scale, and time to at least a two-category improvement), the same competing risk approach will be used as for the primary analysis and results will be summarized by the hazard ratio from the Fine-Gray model.

- Covariate adjustment clarification – for some of the models described below, it might be necessary to combine regions and / or disease severity levels to meet modeling assumptions.

- Unblinding for the final analysis of a study agent:

  Unblinding of study data for final analysis will occur for each test agent independently of other agents, consistent with the master protocol design. That is, once the planned number of recoveries required for a particular agent's comparative analysis are observed, study close-out

procedures for that agent are applied for all study visits and data elements associated with the participants receiving the test agent as well as participants receiving SoC plus placebo during the randomization period for that agent (i.e., concurrently controlled participants) subset to those eligible to receive the agent.

Once the study visits are monitored, data edits are completed, queries are resolved, the database is locked, and treatment assignments are unmasked for that comparative analysis only. Note that, because placebo control participant data are shared across agents, procedures for unmasking data for one agent's final analysis will be established to protect the ability to continue the ongoing study in a double-blinded manner for the remaining agents. It may be that recovery rates are similar enough across all agents in the study to enable a single study close-out with standard procedures. Any necessary modifications or updates to the statistical analysis should be made prior to the study unblinding.

- Handling of Missing Data:

Efforts to minimize loss-to-follow-up will be considerable. However, small amounts of missing data may occur. The plans for handling missing data follow the approach taken in the ACTT-1 study (Beigel, Tomashek et al. 2020). All attempts will be made to collect all data per protocol. Any data point that appears to be erroneous or inexplicable based on clinical judgment will be investigated as a possible outlier. If data points are identified as outliers, sensitivity analyses may be performed to examine the impact of including or excluding the outliers. Any substantive differences in these analyses will be reported.

For time to event outcomes, participants who are lost to follow-up or terminate the study prior to Day 28 and prior to observing/experiencing the event will be censored at the time of their last observed assessment.

For the analyses of the secondary outcomes described in Section 4.3, the following imputation rules will be used for participants who are lost to follow-up, terminate early from the study, or do not have further outcome data available after discharge for any reason:

• Days of Non-invasive ventilation/high-flow oxygen:

  - If the subject's clinical status scale is 3 or 2 at the last observed assessment, then the subject will be considered to be on non-invasive ventilation/high-flow oxygen through

Day 28. The endpoint will be total days when assessments are available plus all imputed days following the last observed assessment.

- If the subject is not on non-invasive ventilation/high-flow oxygen at the last observed assessment, then the subject will be considered to not be on non- invasive ventilation/high-flow oxygen for the remainder of follow-up. Thus, no additional imputed days will be added to the number of days recorded on available assessments.

- Days of ventilation/ECMO:
  - If the subject's clinical status scale is 2 at the last observed assessment, then the subject will be considered to be on ventilation/ECMO through Day 28. The endpoint will be total days when assessments are available plus all imputed days following the last observed assessment.
  - If the subject is not on ventilation/ECMO at the last observed assessment, then the subject will be considered to not be on ventilation/ECMO through Day 28. Thus, no additional imputed days will be added to the number of days recorded on available assessments.

- Days of Oxygen:
  - If the subject's clinical status score is 4, 3, or 2 at the last observed assessment, then the subject will be considered to be on oxygen through Day 28. The endpoint will be total days when assessments are available plus all imputed days following the last observed assessment.
  - If the subject is not on oxygen at the last observed assessment, then the subject will be considered to not be on oxygen through Day 28. Thus, no additional imputed days will be added to the number of days recorded on available assessments.

- Days of Hospitalization
  - If the subject is discharged and no further hospitalization data are available, then the subject will be assumed to not have been readmitted. Thus, no additional imputed days will be added to the number of days recorded on available assessments.

- NEWS - There will be no imputation for incomplete components of the NEWS score.

4.2 ANALYSIS OF THE PRIMARY AND KEY SECONDARY EFFICACY ENDPOINTS

- **Primary Endpoint – Time to Recovery through Day 28**

The primary efficacy analysis for the comparison of each test agent plus SoC versus SoC plus placebo is a stratified test based on the Fine-Gray proportional hazards approach, where stratification is according to geographic region and baseline disease severity on the 8-point ordinal scale (i.e. 2 or 3 or 4 or 5). Other covariates in the model will include age, and sex. The approach is similar to using a stratified log-rank test on time to recovery, providing an estimate of the cumulative incidence of recovery while accounting for the competing risk of death. The primary quantity of interest is the RRR. With no censoring, the hazard rate in each arm can be thought of as the hazard for recovery, treating deaths as never having recovered (Fine and Gray 1999, Gooley, Leisenring et al. 1999). The median time to event and 95% confidence interval will be summarized by treatment arm. Cumulative incidence curves for each treatment arm will be presented, supplemented with the RRR estimate, p-value, along with the number of subjects at risk in each arm at Days 0 (baseline), 2, 4, 7, 10, 14, 21, and 28.

- Sensitivity Analyses

Gray's test on time to recovery will be calculated to obtain the p-value with fewer parametric assumptions (Gray 1988).

As a sensitivity analysis, the main analysis of the primary and two key secondary endpoints (described below) will be repeated in the MITT population.

- Supplemental Analyses

A supplemental analysis will examine the Day 28 binary response of 'recovered' or 'not recovered' which aligns with the 8-point ordinal score at Day 28. Approximately, 95% of study participants are expected to have complete information on this endpoint.

For individuals with a recovery event but missing 8-point ordinal score at Day 28 (~3% of the study population), a multiple imputation approach will be applied to impute those missing values. For each missing observation, 10 imputed observations will be randomly drawn from the subset of individuals with complete data who had a recovery event.

For individuals without a recovery event but missing 8-point ordinal score at Day 28 (~2% of the study population), a multiple imputation approach will be applied to impute those missing values. For each missing observation, 10 imputed observations will be randomly drawn from the subset of individuals with complete data and the same baseline 8-point ordinal score (i.e. 2 or 3 or 4 or 5). Results from the 10 imputed datasets will be combined using Rubin's rule to obtain the point estimate, 95% confidence interval, and p-value.

Those individuals in categories 6-8 would be considered as recovered and those in categories 1-5 are 'not recovered'. The point estimate will be based on a difference in the probability estimates for active vs. control and the 95% CI will be constructed (and p-value obtained) using the Wald method.

Additionally, for each participant, a count of the number of days recovered and out-of-hospital will be computed. The treatment effect will be summarized by the mean difference between groups. This outcome will be analyzed using a linear model with adjustment for geographic region and baseline disease severity on the 8-point ordinal scale (i.e. 2 or 3 or 4 or 5) and an indicator for active vs. placebo treatment. An additional supplemental analysis will analyze this outcome variable using a Wilcoxon-Mann-Whitney test.

Analyses that take into account concomitant medication will be performed. The primary analysis will be repeated, where subjects who take prohibited medications will be treated as treatment failures and will be censored at the time of medication use. Other censoring techniques and additional analyses may be performed.

- **Key Secondary Endpoint #1 - Clinical status (8-point ordinal scale) at Day 14**

The ordinal scale will be used to estimate a proportional odds model by disease strata and region. We will perform a covariate adjusted test to evaluate whether the common odds ratio for treatment is equal to one. The model covariates will include baseline severity (2, 3, 4, or 5), region, sex, and age (continuous). The distribution of severity results will be summarized by treatment group as percentages. The treatment odds ratio, 95% CI, and p-value estimated from the model will be presented.

Supplemental analyses will examine the clinical status using the 8-point ordinal scale at Day 28. This endpoint is expected to be bimodal with individuals more likely to be at the extremes of the scale compared to the Day 14 version.

Analyses of the Day 14 and Day 28 clinical status score will be conducted using multiple imputation. The ~2% of participants who recovered prior to Day 14 but have missing data for the 8-point ordinal scale collected at Day 14 will have 10 observations imputed using randomly selected data from other participants who had complete data for their clinical status (8-point ordinal scale) at Day 14 and a recovery event prior to Day 14.

Participants without a recovery and with missing data for the 8-point ordinal scale at Day 14 (~2% of participants) will have 10 imputed values randomly drawn from the subset of individuals with complete data and the same baseline 8-point ordinal scale. A similar approach will be taken for the Day 28 clinical status score. Results from the 10 imputed datasets will be combined using Rubin's rule.

The number and proportion of subjects along with 95% confidence intervals by category of clinical status will be presented by treatment arm at Days 0 (baseline), 2, 4, 7, 10, 14, 21, and 28. A figure will present stacked bar charts by day with side by side bars for each treatment arm. Histograms will be generated to display the ordinal scale value distributions over time in each treatment group.

- **Key Secondary Endpoint #2 - Mortality through Day 28**

Mortality through Day 28 will be analyzed as a binary endpoint, using a logistic regression model with an indicator variable for the treatment group. Other variables in the regression model will include the 8-point ordinal scale at baseline (2, 3, 4, or 5), region, sex, and age (continuous). The odds ratio from this model will be supplemented with event rates at 14 and 28-days along with a difference between treatment groups and 95% confidence intervals (and p-values) obtained using the Wald method. The number and percentage of subjects who die within 14- and 28-days following randomization will be presented by treatment arm (denominator for the percentages will be the number of subjects in the ITT population in each treatment arm).

Approximately, 95% of study participants are expected to have complete information on this endpoint.

For individuals with a recovery event but missing mortality status at Day 28 (~3% of the study population), a multiple imputation approach will be applied to impute those missing values. For each missing observation, 10 imputed observations will be randomly drawn from the subset of individuals with complete data who had a recovery event.

For individuals without a recovery event but missing mortality status at Day 28 (~2% of the study population), a multiple imputation approach will be applied to impute those missing values. For each missing observation, 10 imputed observations will be randomly drawn from the subset of individuals with complete data and the same 8-point ordinal scale baseline score.

Results from the 10 imputed datasets will be combined using Rubin's rule to obtain the point estimate, 95% confidence interval, and p-value.

- **Tipping Point Sensitivity Analyses for the Primary and Key Secondary Endpoints**

Tipping point sensitivity analyses that systematically and comprehensively vary assumptions about the missing outcomes on the two treatment arms will be applied. These analyses will be two-dimensional and will allow assumptions about the missing outcomes on the two arms to vary independently. These scenarios will include those where dropouts on active drug tend to have worse outcomes than dropouts on placebo. The goal is to explore the plausibility of missing data assumptions under which the conclusions change, i.e., under which there is no longer evidence of efficacy.

Among randomized participants, approximately 5% or participants have missing data for the primary or the two key secondary endpoints (**Table 9**). There are two distinct types of missing data for these endpoints:

1) participants with almost no information (n~42) and

2) those with missing data following a recovery event (n~60).

Among the 1308 randomized participants with a recovery event prior to Day 14, the vast majority (~98%) had complete data for the ordinal score at Day 14 and approximately 99% of those individuals had a score of 7 or 8. Similarly, among the 1527 randomized participants with a recovery event prior to Day 28, only 8 (<1%) died prior to Day 28.

Table 9. Summary of Missing Data by Type for the Primary and Two Key Secondary Endpoints (data as of March 28, 2022)

Endpoint	Total Participants with Missing Data (N=1971)	Post-Recovery Missing Data (N=1971)	Missing Data Prior to a Recovery (N=1971)
Primary Endpoint – Time to Recovery through Day 28	42* (2.1%)	n/a	42 (2.1%)
Key Secondary Endpoint #1 - Clinical status (8-point ordinal scale) at Day 14	77 (3.9%)	35 (1.8%)	42 (2.1%)
Key Secondary Endpoint #2 - Mortality through Day 28	102 (5.2%)	60** (3.0%)	42 (2.1%)

* 5 (11.4%) participants had > 3 days of follow-up and 11 (25%) received any of their randomized study drug

**of the 1527 participants who recovered prior to Day 28, 8 (<1%) were determined to have died prior by Day 28.

To create an imputation strategy that maintains consistency between the primary and two key secondary endpoints, we will first impute the 'mortality status at day 28' endpoint, then the '8-point ordinal scale at 14 days', and finally, the 'time-to-recovery through day 28' primary endpoint. This structure will help us avoid illogical scenarios such as having the same participant with the following imputed values …

- Death at Day 6

- Recovery at Day 10

- Day 14 8-point ordinal scale of 2 (hospitalized, on invasive mechanical ventilation or ECMO)

Clearly, a participant who dies on Day 6 cannot have a recovery on Day 10.

An ordinal logistic regression model will be fitted to the entire ACTIV1 cohort of randomized participants. The outcome variable will be the participant's status at Day 28 and will have three levels. The worst value of the outcome is death, the best value of the outcome is recovery prior to Day 28, and the intermediate value is for those participants who are alive without a recovery by Day 28. The logistic regression imputation model will include covariates for region and severity of illness ordinal score at the time of randomization. A parameter, Θ, will vary from 0.1 to 10 and correspond to an odds ratio parameter for the ordinal logistic model. Values of Θ less than 1.0 will be associated with better than average outcomes and values great than 1.0 will have worse than average outcomes. Specifically, for each level of the ordered response, the value of $\log(\Theta)$ will be added to the logit function with the covariates associated with the participants with missing data. Then a random uniform variable will be

generated and compared to the imputed probability distribution. With the imputed outcome of either 'death', 'recovery', or 'alive without recovery', the time-to-recovery and the Day 14 8-point ordinal scale response will be imputed by selecting a participant at random with the same severity of illness at randomization, region, and corresponding status at Day 28 (e.g. 'death', 'recovery', or 'alive without recovery').

On average, a value of $\Theta = 0.1$ will correspond to an expected Day 28 mortality of < 1% whereas a value of $\Theta = 10$ will correspond to an expected Day 28 mortality of ~ 49% (see **Table 10**). As special cases, values are $\Theta = 0$ and $\Theta = \infty$ will be used to denote the special cases of 'best possible outcome' and 'worst possible outcome', respectively. As an example, for a participant who withdrew on Day 3, the best possible outcome (i.e. $\Theta = 0$) would correspond to: Recovery at 3, Day 14 8-point ordinal scale of 8, and Alive at Day 28. Similarly, for a participant who withdrew on Day 3, the worst possible outcome (i.e. $\Theta = \infty$) would correspond to: Death at Day 3 and Day 14 8-point ordinal scale of 1.

Table 10. Probability, on average, of recovery and death by Day 28 for various parameter settings

Parameter Setting	Recovery through Day 28	Death through Day 28
$\Theta = 0$ "Best Case"	100%	0%
$\Theta = 0.1$	98%*	1%
$\Theta = 0.2$	95%	2%
$\Theta = 0.5$	90%	6%
$\Theta = 1$ "Missing at Random"	82%	10%
$\Theta = 2$	71%	18%
$\Theta = 5$	51%	34%
$\Theta = 10$	35%	49%
$\Theta = \infty$ "Worst Case"	0%	100%

*the exact values will depend on the participant's covariates

For each parameter setting, 10 imputed datasets will be constructed for each individual with missing data for the primary endpoint. For each imputed dataset, the primary and two key secondary analyses will be re-run and the results of the 10 models will be combined using Rubin's rules. The results of the tipping point analysis will allow for a comparison of point estimates and p-values across the various values of Θ (see **Table 11** for an example).

Table 11 will be repeated for the two key secondary endpoints, and this analysis will be repeated for the comparison of each study drug vs. the corresponding placebo. Separate models will be constructed for participants with incomplete data following a recovery event (e.g. withdrawal of consent after discharge from the hospital).

Table 11. Example summary of the tipping point analysis for the time-to-recovery endpoint

RRR* 95% CI p-value						Active				
		Θ = 0	Θ = 0.1	Θ = 0.2	Θ = 0.5	Θ = 1.0	Θ = 2	Θ = 5	Θ = 10	Θ = ∞
	Θ=0									
	Θ=0.1									
	Θ=0.2									
	Θ=0.5									
Placebo	Θ=1									
	Θ=2									
	Θ=5									
	Θ=10									
	Θ=∞									

*the values within each cell will correspond to the estimated RRR, 95% CI, and p-value for the particular parameter setting (e.g. $\Theta_{Placebo}=1$, $\Theta_{Active}=10$).

The above description of the tipping point analysis has focused on the missing data of the subset of participants with no recovery. For the two key secondary endpoints, it is expected that 2-3% of participants will have some missing data following a recovery event.

For individuals with a recovery event but missing mortality status at Day 28 (~3% of the study population) or the Day 14 clinical status score, a multiple imputation approach will be applied to impute those missing values. For each missing observation, 10 imputed observations will be randomly drawn from the subset of individuals with complete data who had a recovery event, and the imputed datasets will be combined as described above.

4.3 ANALYSIS OF THE OTHER SECONDARY EFFICACY ENDPOINTS

- Time to improvement of one and two categories in ordinal scale

Two separate endpoints are considered here. The first endpoint will examine the time to a one category improvement in the 8-point ordinal scale. The second endpoint will examine the time to a two category improvement in the 8-point ordinal scale. These time-to-event endpoints will be computed through Day 28 will also be analyzed using the same Fine-Gray proportional hazards approach as for the primary efficacy endpoint and stratification using the randomization factors. Proportions of one and two category improvements at 14 and 28-days along with 95% confidence intervals overall for each treatment arm along with a difference between treatment arms and 95% confidence intervals obtained

using the Miettinen-Nurminen method. Calculations of these endpoints will rely on the 8-point ordinal scale that is collected on a daily basis while the participant is hospitalized.

- Mean change in ordinal scale at specific time points

Change in clinical status scale from baseline at specific time points will be summarized by proportions (e.g., proportion who have a 1-, 2-, 3-, or 4+ point improvement or 1-, 2-, 3-, 4-point worsening). A table will present the proportion of people on Days 2, 4, 7, 10, 14, 21, and 28 within each category of change by treatment arm along with 95% confidence intervals. The mean change in the ordinal scale difference between treatment arms along with 95% confidence interval will also be reported. Additionally, side-by-side histograms (of an agent versus control) of the proportions with various changes will be presented.

- Differences in total severity score (TSS) over time

Differences in total severity score (TSS) over time will assessed by fitting a longitudinal Markov model to a modified version of the 8-point clinical scale value on a daily basis between days 1 and 28. The average difference across days in study (analogous to the area under the severity curve) will be estimated with 95% confidence limits in the context of the longitudinal model. The longitudinal mixed model will be developed as follows. The value of the ordinal scale on day t will be assumed to depend on the patient's prior outcome trajectory only through the prior day's value, and treatment group assignment. The general form of the model will be:

$$\Pr\left(Y_{it} \le k \middle| H_{i,t-1}, Y_{i.t-1} = x, Z_i = z, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, e_i\right) \equiv \pi_{itk|xz},$$

$$\log \frac{\pi_{itk|xz}}{1-\pi_{itk|xz}} = \alpha_{kx} + t\beta_1 + t^2\beta_2 + z(\gamma_0 + t\gamma_1 + t^2\gamma_2),$$

where Y_{it} is outcome of the i-th patient on day t, $H_{i,t-1} = (Y_{i1}, \dots, Y_{i,t-1})$ is the patient's outcome trajectory through day $t-1$, Z_i is a binary indicator of treatment group assignment (1=active agent plus SoC, 0=placebo plus SoC), and the symbols $\alpha_{kx}, \beta_1, \beta_2, \gamma_0, \gamma_1\gamma_2$ represent unknown fixed effect parameters to be estimated from the data. The model assumes that on each day, the odds of an outcome in one of the k lowest categories depends jointly on treatment assignment, days since randomization, and *x* is the patient's prior day outcome category. Parameters will be estimated by maximum likelihood using 1 record per patient per day from the first day post-randomization until the earlier of day 28, death or loss to follow-up. Days between death and day 28 do not contribute to the likelihood because death is an absorbing state and hence $\alpha_{11} = \infty$ and $\alpha_{21} = \alpha_{31} \cdots = \alpha_{81} = -\infty$. The estimation procedure will assume that loss to follow-up is conditionally ignorable given treatment assignment and the patient's current observed outcome trajectory. A feature of the trial's outcome assessment schedule is that only patients who are hospitalized have their outcome assessed on days

other than 7, 10, 14, 21, and 28. If a patient is not hospitalized, the patient's outcome category will be at least 7 but the exact value (7 versus 8) will only be collected on days 7, 10, 14, 21, and 28. To mitigate potential bias from informatively missing data, categories 7 and 8 will be combined into a single category for this analysis. This modification guarantees that a patient will have non-missing data on all days 1 through 28 if the patient's vital status and hospitalization status is known on each day and the patient's outcome is assessed per protocol on each day that the patient is alive and in-hospital.

- Days alive and free of supplemental oxygen

Days alive and free of supplemental oxygen and its components, duration of oxygenation days and total days alive, will be summarized in a table using mean (median) and standard deviations (quartiles) by treatment arm. Point estimates for the difference between active and placebo for days alive and free of supplemental oxygen and associated 95% confidence intervals will be obtained using a linear model with treatment group and baseline 8-point ordinal score (i.e. 2 or 3 or 4 or 5) as categorical predictors. The duration of oxygenation days will be computed using the 8-point ordinal scale that is collected each day the participant is hospitalized. If a participant is rehospitalized after an initial recovery and the 8-point ordinal scale is not available, the participant will be assumed to have an 8-point ordinal score value of 2 for all days of the rehospitalization.

The incidence of new oxygen use will be analyzed by treatment arm. This will only include subjects in ordinal scale category 5 at enrollment (i.e. 5: Hospitalized, not requiring supplemental oxygen – requiring ongoing medical in-patient care). New use will be identified by a post-enrollment score of 4 or 3 or 2. The number and percent of subjects reporting new use and 95% confidence interval will be reported. Comparisons between treatment arms will be presented as differences in proportions with 95% confidence intervals using the Miettinen-Nurminen method.

- Days alive and free of non-invasive ventilation/high flow oxygen

Days alive and free of non-invasive ventilation/high flow oxygen and its components, duration of non-invasive ventilation/high flow oxygen days will be summarized in a table using mean (median) and standard deviations (quartiles) by treatment arm. Point estimates and associated 95% confidence intervals will be obtained using a linear model with treatment group and baseline 8-point ordinal score (i.e. 2 or 3 or 4 or 5) as categorical predictors. The duration of non-invasive ventilation/high flow oxygen days will be computed using the 8-point ordinal scale that is collected each day the participant is hospitalized. If a participant is rehospitalized after an initial recovery and the 8-point ordinal scale is not available, the participant will be assumed to have an 8-point ordinal score value of 2 for all days of the rehospitalization.

The incidence of new Non-Invasive Ventilation/High Flow Oxygen use will be analyzed by treatment arm. This will only include subjects in category 4 or 5 at enrollment. New use will be identified by a post-

enrollment score of 3 or 2. Comparisons between treatment arms will be presented as differences in proportions with 95% confidence intervals using the Miettinen-Nurminen method.

- Days alive and free of invasive mechanical ventilation/ECMO

Days alive and free of invasive mechanical ventilation / ECMO and its components, total days alive and duration of invasive mechanical ventilation/ECMO days, will be summarized in a table using mean (median) and standard deviations (quartiles) by treatment arm. Point estimates for differences in the days alive and free of invasive mechanical ventilation / ECMO and associated 95% confidence intervals will be obtained using a linear model with treatment group and baseline 8-point ordinal score (i.e. 2 or 3 or 4 or 5) as categorical predictors. The duration of invasive mechanical ventilation/ECMO days will be computed using the 8-point ordinal scale that is collected each day the participant is hospitalized. If a participant is rehospitalized after an initial recovery and the 8-point ordinal scale is not available, the participant will be assumed to have an 8-point ordinal score value of 2 for all days of the rehospitalization.

The incidence of new Invasive Mechanical Ventilation/ECMO use will be analyzed by treatment arm. This will only include subjects in category 5, 4, or 3 at enrollment. New use will be identified by a post-enrollment score of 2. The number and percent of subjects reporting new use and the 95% confidence interval will be reported. Comparisons between treatment arms will be presented as differences in proportions with 95% confidence intervals using the Miettinen-Nurminen method.

- Days alive and out of hospitalization

Days alive and out of hospital along with its components, duration of hospitalizations and total days alive, will be summarized in a table using mean (median) and standard deviations (quartiles) by treatment arm. Point estimates for the difference between active and placebo agents for days alive and out of hospital and associated 95% confidence intervals will be obtained using a linear model with treatment group and baseline 8-point ordinal score (i.e. 2 or 3 or 4 or 5) as categorical predictors.

4.4 SAFETY ANALYSES

Safety endpoints include death through Day 28 (as assessed at Day 29), SAEs, and Grade 3 and 4 AEs. These events will be analyzed individually and as a composite endpoint. The time to this composite endpoint will be defined as the elapsed time (in days) from baseline to the earliest date of any of the events. Any subjects that are lost to follow-up or terminated early prior to experiencing any of the events will be censored at the day of their last observed assessment. Subjects who complete follow-up

but do not experience any of the events will be censored at the day of their Day 29 visit. Time-to-event methods will be used for death and the composite endpoint. A table will present the hazard ratio estimate with 95% confidence intervals and log rank p-values. Differences in time-to-event endpoints by treatment will be summarized with cumulative incidence curves. In addition, risk differences between treatment groups and 95% CIs using the Miettinen-Nurminen method will be estimated for each type of event.

Similar analyses will be conducted for the following AEs:

- Any Grade 2 suspected drug-related hypersensitivity reactions associated with study product administration will be reported as an AE.
- Any newly diagnosed infection (other than COVID-19) at any time during the study.
- Any AEs leading to dose modification
- Any AEs leading to discontinuation from the study.

AEs will be coded in accordance with the current version of the Medical Dictionary for Regulatory Activities (MedDRA). AEs will be graded with regard to severity according to criteria defined in the Common Terminology Criteria for Adverse Events version 5.0 (CTCAE) and relationship to COVID-19 or study intervention. Each AE will be counted once only for a given participant. AEs will be presented by system organ class, preferred term (PT), duration (in days), start- and stop-date. AEs leading to premature discontinuation from the study intervention and serious AEs will be presented in a listing.

4.4.1 LABORATORY DATA

Clinical safety laboratory adverse events are collected Day 1, 3, 5, 8, 11 and Day 15 and 29 if able to return to clinic or still hospitalized. Parameters evaluated include white blood cell count, differential, Hgb, PLT, creatinine, glucose, total bilirubin, AST, ALT, ALP, PT or INR, d-dimer, cardiac troponin, and C-reactive protein.

Descriptive statistics for baseline and post-baseline time points will be generated. Additionally, shift tables will be presented. The shift tables tabulate the number of lab values determined to be “Normal” and “Abnormal” at baseline and post-baseline time points. The classifications on shift tables will categorize lab values as “Normal” when they are in acceptable range. Laboratory values outside the normal range will be graded on the Labs Listing as >1x, >2x, >3x, >5x, or >10x the upper limit (or <0.1x, <0.5x, <1x the lower limit) of normal value based on appropriate increases or decreases. Laboratory data will also be summarized graphically to show the magnitude and changes in individual subject values over time relative to normal ranges. Details of any abnormalities will be listed.

4.4.2 VITAL SIGNS

Vital sign measurements include pulse, systolic blood pressure, respiratory rate, SpO2 and oral temperature.

4.4.3 PRIOR AND COMCOMITANT MEDICATIONS

Concomitant medications will be coded to the Anatomical Therapeutic Classification using the WHO Drug Dictionary. The use of prior, concomitant medications taken, and changes in concomitant medications during the study will be recorded on the CRFs. The use of concomitant medications during the study will be summarized by ATC1, ATC2 code, and treatment group for the ITT population. The summaries will be repeated for each test agent.

4.4.4 TREATMENT COMPLIANCE

Treatment compliance will be assessed based upon the administration rules for each of the test agents. For treatments that are one-time infusions (e.g. infliximab or abatacept), compliance will be summarized by the percentage of participants starting the infusion, the mean duration of the infusion, and the total amount of drug received (mg). For oral medications (e.g. cenicriviroc), compliance will be summarized by the percentage of participants taking the loading dose and starting the maintenance dose, mean number of maintenance doses, and mean duration (days) of the oral medication.

4.4.5 PREGNANCY TESTS

A set of listings of pregnancies and outcomes will be presented.

4.5 EXPLORATORY ANALYSES

The probability of falling into category “i” or better will be compared between arms for each baseline severity score category. Exploratory analyses will also compare the difference between each test agent plus SoC and SoC plus placebo in the 8-point ordinal scale at each study day for which the clinical scale is administered (in-hospital and at Days 14, 21, and 28).

A set of Bayesian analyses will be conducted for the primary endpoint using informative, enthusiastic, and skeptical priors as described in Wijeysundera et al. (2009) and Zampieri et al. (2021). Three different prior distributions will be assumed for the odds ratio parameter for the treatment effect in the proportional odds model. For the informative prior distribution, the log odds ratio (active drug vs. placebo) will have a mean of 0 and a standard deviation of 10 on the log-odds scale. For the skeptical prior, the log odds ratio will have a mean of 0 and a standard deviation for the log-odds ratio that provides 5% probability for a response that is better than the assumed alternative hypothesis (e.g. OR=1.25). For the enthusiastic prior, the assumed log-odds ratio will use the treatment effect assumed for the estimation of the sample size (e.g. OR~1.25) and a standard deviation that provides 5% probability of no benefit (e.g. OR<=1.00). Similar analyses will be conducted for the 28-day mortality binary endpoint. For these analyses, the enthusiastic prior will correspond to an odds ratio < 1.0 indicating treatment benefit for the active drug.

A separate statistical analysis plan will be created for the analyses of biomarkers.

4.6 BASELINE DESCRIPTIVE STATISTICS

Baseline characteristics will be summarized by test agent versus control for each agent and study population. For continuous measures the mean and standard deviation will be summarized. Categorical variables will be described by the proportion in each category (with the corresponding sample size). Baseline characteristics will include age, sex, race, ethnicity, symptom duration, hypertension, obesity, diabetes, coronary disease, history of heart failure, history of cancer, asthma, COPD, chronic pulmonary disease, tuberculosis, HIV, severe liver disease, severe kidney disease, days since hospitalization, and baseline 8-point ordinal score.

4.7 SUB-GROUP ANALYSES

The following subgroups have been pre-specified for the primary outcome and the two key secondary outcomes:

- Geographic region
- Duration of symptoms prior to randomization (<= Median; > Median)
- Duration of symptoms prior to randomization (based on quartiles)
- Hospitalization duration prior to randomization (<= Median, > Median)
- Baseline disease severity based on the 8-point ordinal scale

- Age (<40; 40-64; 65 and older)
- Race (White; Black/African American; Asian; Other)
- Ethnicity (Hispanic, non-Hispanic)
- Sex (Female; Male)
- Comorbidities: 0, 1, 2, 3+
- Baseline use of dexamethasone
- Co-enrollment in any ACTIV trial prior to randomization in ACTIV-1 IM
- Individual comorbidities: hypertension, obesity, diabetes, coronary disease, history of heart failure, history of cancer, asthma, COPD, chronic pulmonary disease, severe liver disease, severe kidney disease

Any subgroup that has a frequency <5% will not be generated. Subgroups with frequencies <10% may also have issues with model convergence that dictate that they cannot be generated.

The primary analysis will also be repeated for the primary endpoint and the two key secondary endpoints using the subgroups defined above. Each subgroup will be considered separately and the tabular and graphical summaries described in the previous section will be replicated for each subgroup. In addition, a forest plot will be generated to display the overall treatment hazard ratio (or RRR) estimate and CI from each of the within-stratum analyses. These analyses will be performed in the ITT and MITT populations.

Interaction tests will be conducted to determine whether the effect of treatment varies by subgroup.

In addition to the subgroup analyses described above, primary and key secondary and safety analyses will be performed within additional subgroups as outlined in **Appendix 2**.

4.8 REFERENCES

Beigel, J. H., K. M. Tomashek, L. E. Dodd, A. K. Mehta, B. S. Zingman, A. C. Kalil, E. Hohmann, H. Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R. W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T. F. Patterson, R. Paredes, D. A. Sweeney, W. R. Short, G. Touloumi, D. C. Lye, N. Ohmagari, M. D. Oh, G. M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M. G. Kortepeter, R. L. Atmar, C. B. Creech, J. Lundgren, A. G. Babiker, S. Pett, J. D. Neaton, T. H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak and H. C. Lane (2020). "Remdesivir for the Treatment of Covid-19 - Preliminary Report." N Engl J Med.

Cao, B., Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, X. Li, J. Xia, N. Chen, J. Xiang, T. Yu, T. Bai, X. Xie, L. Zhang, C. Li, Y. Yuan, H. Chen, H. Li, H. Huang, S. Tu, F. Gong, Y. Liu, Y. Wei, C. Dong, F. Zhou, X. Gu, J. Xu, Z. Liu, Y. Zhang, H. Li, L. Shang, K. Wang, K. Li, X. Zhou, X. Dong, Z. Qu, S. Lu, X. Hu, S. Ruan, S. Luo, J. Wu, L. Peng, F. Cheng, L. Pan, J. Zou, C. Jia, J. Wang, X. Liu, S. Wang, X. Wu, Q. Ge, J. He, H. Zhan, F. Qiu, L. Guo, C. Huang, T. Jaki, F. G. Hayden, P. W. Horby, D. Zhang and C. Wang (2020). "A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19." N Engl J Med 382(19): 1787-1799.

Fine, J. P. and R. J. Gray (1999). "A Proportional Hazards Model for the Subdistribution of a Competing Risk." Journal of the American Statistical Association 94(446): 496-509.

Gooley, T. A., W. Leisenring, J. Crowley and B. E. Storer (1999). "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators." Stat Med 18(6): 695-706.

Gray RJ (1988). “A class of K-sample tests for comparing the cumulative incidence of a competing risk.” Ann Stat 16:1141-1154.

Holm SA. (1979) “A simple sequentially rejective multiple test procedure.” Scandanavian Journal of Statistics; 6: 65-70.

Huang, C., Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang and B. Cao (2020). "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." Lancet 395(10223): 497-506.

Hung J, Wang S-J, O'Neill R (2007). “Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials.” J Biopharm Stat; 17(6):1201-10.

Hwang IK, Shih WJ, J S De Cani JS (1990). “Group sequential designs using a family of type I error probability spending functions” Stat Med; 9(12):1439-45.

Smith, G. B., D. R. Prytherch, S. Jarvis, C. Kovacs, P. Meredith, P. E. Schmidt and J. Briggs (2016). "A Comparison of the Ability of the Physiologic Components of Medical Emergency Team Criteria and the

U.K. National Early Warning Score to Discriminate Patients at Risk of a Range of Adverse Clinical Outcomes." Crit Care Med 44(12): 2171-2181.

Tamhane AC, Mehta CR, Liu L (2010) “Testing a Primary and a Secondary Endpoint in a Group Sequential Design.“ Biometrics; 66(4): 1174–1184.

Whitehead J. (1993) “Sample Size Calculations for Ordered Categorical Data.” Stat Med; vol. 12, 2257-2271.

Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A (2009). “Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials.” J Clin Epidemiol; 62: 13-21 e15.

Zampieri FG, Casey JD, Shankar-Hari M; Harrell Jr. FE, and Harhay MO (2021). “Using Bayesian Methods to Augment the Interpretation of Critical Care Trials: An Overview of Theory and Example Reanalysis of the Alveolar Recruitment Trial.” Am J Resp Crit Care Med. Am J Respir Crit Care Med; 203(5): 543-552.

4.9 APPENDICES

APPENDIX 1. ACTIV-1 Roster of Statisticians

Team	Name	Email	Attends DSMB Meetings
Unblinded Statistical Team	**Michael Proschan**	proscham@niaid.nih.gov	**X**
	Tyler Bonnet	tbonnett@uw.edu	**X**
	Jyotsna Garg	jyotsna.garg@duke.edu	**X**
	Eric Yow	eric.yow@duke.edu	**X**
	Zhen Huang	zhen.huang@duke.edu	
	Hwasoon Kim	hwasoon.kim@duke.edu	
	Hussein Al-Khalidi	hussein.al-khalidi@duke.edu	**X**
	Sean O'Brien	sean.m.obrien@duke.edu	**X**
Blinded Statistical Team	**Lisa LaVange**	lisa.lavange@unc.edu	**X**
	Steven McNulty	steven.mcnulty@duke.edu	**X**
	Adam Silverstein	adam.silverstein@duke.edu	
	Jun Wen	jun.wen@duke.edu	
	Pearl Zakroysky	pearl.zakroysky@duke.edu	**X**
	Susan Halabi	susan.halabi@duke.edu	**X**
	Kevin Anstrom	kevin.anstrom@unc.edu	**X**

APPENDIX 2. Additional Tables and Listings

Table	Demographic and Baseline Characteristics – US	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics – Latin America	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region - Argentina	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region - Brazil	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region - Mexico	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region - Peru	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region – USA – Northeast	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region – USA – Midwest	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region – USA – South	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics by Geographic Region – USA – West	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics – Moderate Disease Severity at Baseline	Intent-to-Treat Population
Table	Demographic and Baseline Characteristics – Severe Disease Severity at Baseline	Intent-to-Treat Population
Table	Time to Recovery by US vs. Latin America	Intent-to-Treat Population
Table	Time to Recovery by Disease Severity at Baseline (Moderate vs. Severe)	Intent-to-Treat Population
Table	Time to Recovery by Remdesivir Use at Randomization	Intent-to-Treat Population
Table	Time to Recovery by Steroid Use at Randomization	Intent-to-Treat Population
Table	Time to Recovery by Immune Modulator Use at Randomization	Intent-to-Treat Population
Table	Time to Recovery by Anticoagulant Use at Randomization	Intent-to-Treat Population
Table	Time to Recovery by Baseline CRP (<=75 mg/L; >75 mg/L)	Intent-to-Treat Population
Table	Proportional Odds Model of Clinical Status Score at Day 14 by US vs. Latin America	Intent-to-Treat Population
Table	Proportional Odds Model of Clinical Status Score at Day 14 by Disease Severity at Baseline (Moderate vs. Severe)	Intent-to-Treat Population

Table	Proportional Odds Model of Clinical Status Score at Day 28 by US vs. Latin America	Intent-to-Treat Population
Table	Proportional Odds Model of Clinical Status Score at Day 28 by Geographic Region	Intent-to-Treat Population
Table	Proportional Odds Model of Clinical Status Score at Day 28 by Disease Severity at Baseline (Moderate vs. Severe)	Intent-to-Treat Population
Table	28-Day Mortality by US vs. Latin America	Intent-to-Treat Population
Table	28-Day Mortality by Disease Severity at Baseline (Moderate vs. Severe)	Intent-to-Treat Population
Table	28-Day Mortality by Remdesivir Use at Randomization	Intent-to-Treat Population
Table	28-Day Mortality by Steroid Use at Randomization	Intent-to-Treat Population
Table	28-Day Mortality by Immune Modulator Use at Randomization	Intent-to-Treat Population
Table	28-Day Mortality by Anticoagulant Use at Randomization	Intent-to-Treat Population
Table	28-Day Mortality by Baseline CRP (<=75 mg/L; >75 mg/L)	Intent-to-Treat Population
Table	14-Day and 28-Day Mortality Rates by US vs. Latin America	Intent-to-Treat Population
Table	14-Day and 28-Day Mortality Rates by Geographic Region	Intent-to-Treat Population
Table	14-Day and 28-Day Mortality Rates by Disease Severity at Baseline (Moderate vs. Severe)	Intent-to-Treat Population
Table	14-Day and 28-Day Mortality Rates by Baseline CRP (<=75 mg/L; >75 mg/L)	Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 - US	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 – Latin America	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region - Argentina	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region – Brazil	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region – Mexico	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region – Peru	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events	Modified Intent-to-Treat

	through Day 28 by Geographic Region – USA – Northeast	Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region – USA – Midwest	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region – USA – South	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 by Geographic Region – USA – West	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 – Moderate Disease Severity at Baseline	Modified Intent-to-Treat Population
Table	Overview of Safety Composite, Individual Endpoints, and Adverse Events through Day 28 – Severe Disease Severity at Baseline	Modified Intent-to-Treat Population
Table	Most Common Adverse Events through Day 28 Occurring in >10% of Infliximab Subjects	Modified Intent-to-Treat Population
Table	Most Common Adverse Events through Day 60 Occurring in >10% of Infliximab Subjects	Modified Intent-to-Treat Population
Listing	Prior Medications	Intent-to-Treat Population
Listing	Concomitant Medications	Intent-to-Treat Population