Statistical Analysis Plan

This document applies to both substudies

A Multicenter, Adaptive, Randomized, Blinded Controlled Trial of the Safety and Efficacy of Investigational Therapeutics for Hospitalized Patients with Acute Respiratory Distress Syndrome Associated with COVID-19

Version 1.0

05 August 2021

NCT06729606

(Master protocol: NCT04843761)

Statistical Analysis Plan

Version 1.0

Therapeutics for Severely Ill Inpatients with COVID-19 (TESICO)

A Multicenter, Adaptive, Randomized, Blinded Controlled Trial of the Safety and Efficacy of Investigational Therapeutics for Hospitalized Patients with Acute Respiratory Distress Syndrome Associated with COVID-19

Clinicaltrials.gov identifier: NCT04843761
EudraCT number: 2021-001650-56

Version	Date	Who	Comments
1.0	05 August 2021	BG	TESICO ACTIV-3b INSIGHT 015 Protocol v2.0, August 2021, Investigational agents: aviptadil and remdesivir

Contents

1 Introduction 4
1.1 Objective of the Statistical Analysis Plan 4
1.2 Description of the Study Design 4
1.3 Randomization 8
1.4 Sample Size Estimates 8
2 Interim DSMB Reviews: Goals and Format 9
3 Analysis Principles 12
4 Enrollment and Eligibility 17
5 Baseline Characteristics 17
6 Administration of Study Treatment 20
6.1 Infusion of Aviptadil (Active or Placebo) 20
6.2 Infusion of Remdesivir (Active or Placebo) 24
7 Safety Analyses 25
7.1 Safety Analyses for the Aviptadil and Remdesivir Cohorts 26
7.2 Additional Safety Analyses for the Aviptadil Cohort 29
8 Efficacy Analyses 30
8.1 Primary Efficacy Endpoint 30
8.2 Primary Analysis 31
8.3 Key Secondary Outcomes 33
8.4 Other Secondary Outcomes 35
8.5 Subgroup Analyses 36
9 Assessment of Independence in the Factorial Cohort 38
10 Interim Monitoring Guidelines for the DSMB 39
10.1 Early Assessment of Safety 39
10.2 Interim Monitoring Boundaries 39
10.3 Interim Monitoring for Futility 40
11 Data Completeness and Study Conduct 40
12 SARS-CoV-2 Antigen, Antibody, and RNA Levels 41
13 Exploratory Analyses 42
13.1 Disease Progression Risk Score 42
14 Unblinding of Treatment Comparisons 42
15 Distribution of Reports 42
16 References 44
Appendix A. Categories of the Primary Outcome (Recovery) 45
Appendix B. Definition of Clinical Organ Failure and Serious Infection 46

Appendix C. Protocol-specified Exempt Serious Events (PSESE) 47
Appendix D. Safety Data Collection 48
Appendix E. Schedule of Assessments 51
Appendix F. List of Acronyms 54

1 Introduction

1.1 Objective of the Statistical Analysis Plan

The objective of this statistical analysis plan (SAP) is to provide a description of the general analytic strategy and the statistical methods that will be used to analyze the data for the TESICO (Therapeutics for Severely Ill Inpatients with COVID-19) protocol. The goal of TESICO is to evaluate the safety and efficacy of investigational agents aimed at improving outcomes for patients with acute respiratory failure related to COVID-19. TESICO is a sister protocol to the TICO master protocol, with focus on patients with critical respiratory failure (i.e., those receiving high-flow nasal cannula, non-invasive ventilation, invasive mechanical ventilation or ECMO to treat acute hypoxemic respiratory failure caused by SARS-CoV-2 pneumonia).

The master protocol is for a phase III randomized, blinded, placebo-controlled platform trial that allows investigational agents to be added and dropped during the course of the study for efficient testing of new agents against control within the same trial infrastructure. When more than one agent is being tested concurrently, a factorial design may be employed, or participants may be randomly allocated in parallel arms.

This version of the SAP describes statistical analyses for the first two investigational products, Vasoactive Intestinal Peptide (VIP; aviptadil is the generic name for the synthetic peptide, developed by NeuroRX, Inc.) and remdesivir (Gilead Sciences, Inc.). All participants receive standard of care (SOC) and will be randomized to receive one or two investigational agents or matching placebo in addition to SOC, as described in section 1.2.

This SAP:

- Provides a short description of the study design (sections 1.2-1.4)
- Describes goals of the interim reviews by the independent DSMB and the planned format of the review meetings (section 2)
- Describes the planned data analyses presented in the reports to the DSMB (sections 3-13). General analysis principles are summarized in section 3, safety analyses are described in section 7, efficacy analyses in section 8, and interim monitoring guidelines in section 10.
- Describes data summaries to be provided regularly to study leadership to aid in monitoring trial conduct and data quality; these data summaries will be pooled across treatment groups, and will be restricted to enrollment, baseline data, and summaries of data completeness and study conduct.

The SAP for TESICO will be updated by the blinded study statisticians prior to unblinding. It may also be updated based on protocol amendments.

1.2 Description of the Study Design

This section is adapted from Section 1 and Appendix H1 of the TESICO protocol version 2.0.

Design

TESICO (Therapeutics for Severely Ill Inpatients with COVID-19) is a master protocol to evaluate the safety and efficacy of investigational agents aimed at improving outcomes for patients with critical acute respiratory failure caused by SARS-CoV-2 pneumonia.

The protocol is for a phase III randomized, blinded, controlled platform trial that allows investigational agents to be added and dropped during the course of the study for efficient testing of new agents against control within the same trial infrastructure.

In this section, we are describing the trial design for the first two investigational products, aviptadil and remdesivir. In short, the trial consists of a 2x2 factorial for aviptadil versus matched placebo and remdesivir versus matched placebo, and participants who are not eligible to be randomized in the factorial will be randomized 1:1 to one of two treatment groups, either aviptadil versus matched placebo, or remdesivir versus matched placebo, depending on eligibility. All participants receive standard of care (SOC), plus the randomized treatment assignment. Corticosteroid therapy is recommended as part of SOC for all participants, unless contraindicated.

Specifically, the trial includes four strata of participants (referred to as "**design strata**") (Figure 1 on the next page):

Stratum 1: Participants who are eligible for aviptadil and remdesivir, and have not received any remdesivir prior to randomization. These participants will be randomized in a 2x2 factorial to the four possible combinations of aviptadil, remdesivir, and the matching placebos for these drugs: 1) aviptadil + remdesivir placebo; 2) aviptadil placebo + remdesivir; 3) aviptadil + remdesivir; and 4) aviptadil placebo + remdesivir placebo.

Stratum 2: Participants who are not eligible to receive remdesivir (contraindication). These participants will be randomized to aviptadil versus placebo only.

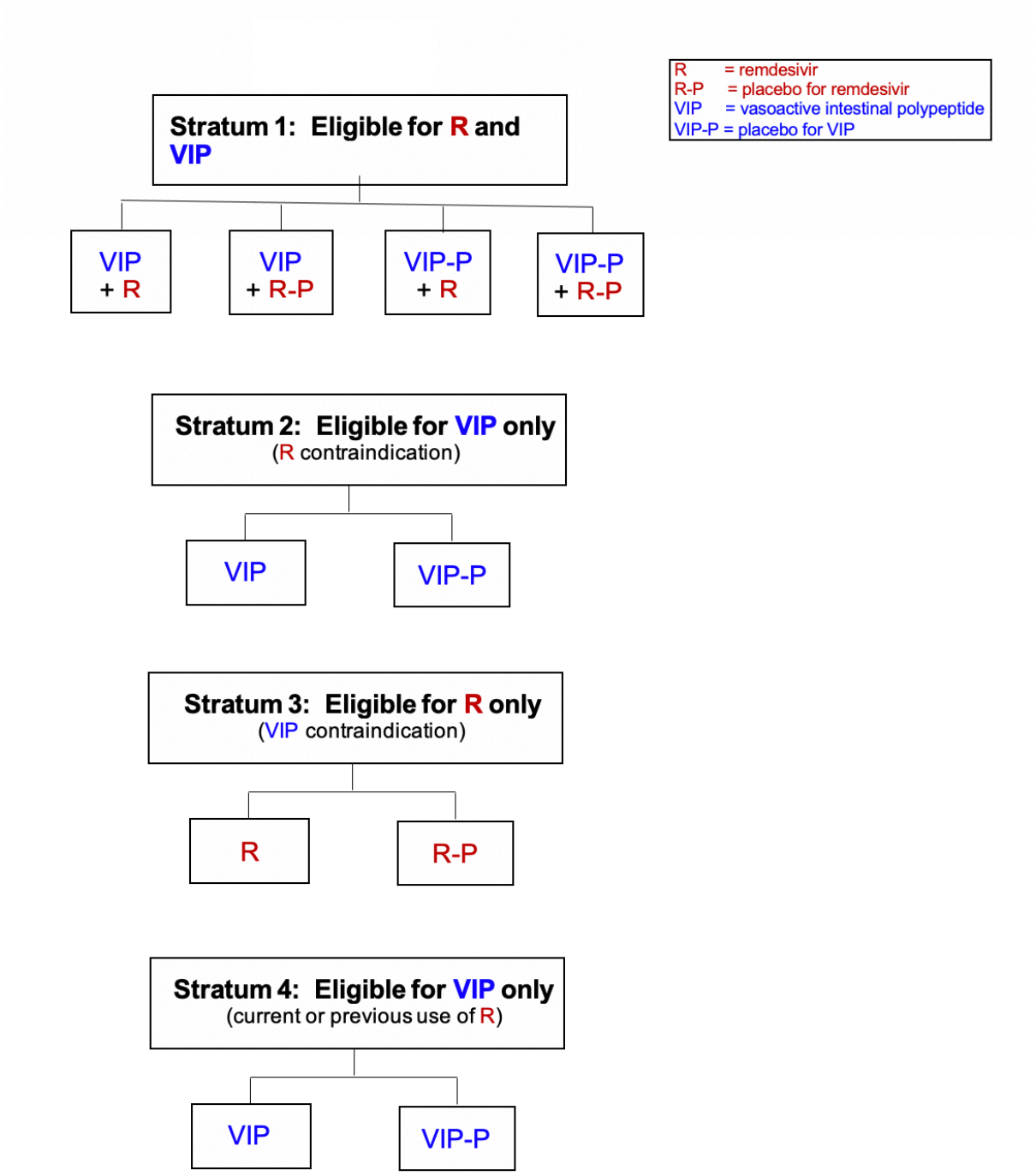
Stratum 3: Participants who are not eligible to receive aviptadil (contraindication). These participants will be randomized to remdesivir versus placebo only.

Stratum 4: Participants who have received remdesivir prior to randomization, and are eligible for aviptadil. These participants will be randomized to aviptadil versus placebo only.

Each randomization of investigational agent versus placebo will use a 1:1 allocation, and equal allocation to the four treatment combinations in the 2x2 factorial. Statistical analyses will compare aviptadil versus placebo, pooling participants in strata 1, 2, and 4, and will compare remdesivir versus placebo, pooling participants in strata 1 and 3. To achieve this, it is estimated that 800 participants will have to be enrolled (640 for each pairwise comparison: aviptadil vs. placebo, and remdesivir vs. placebo).

The design assumes that the effects of aviptadil and remdesivir are independent of each other; a possible interaction between the two investigational agents will be assessed in the 2x2 factorial, although power is limited for the interaction test.

Figure 1: Study Design of TESICO. The study includes 4 "design strata": a 2x2 factorial for aviptadil (VIP) versus placebo and remdesivir versus placebo, and 3 strata with 1:1 randomizations to either investigational agent versus placebo.



Stratum	Percent of Patients
1	60
2	10
3	10
4	20

Sample size for VIP = strata 1, 2 and 4
Sample size for R = strata 1 and 3

Population
The study population consists of inpatient adults (≥18 years) who have documented SARS-CoV-2 infection within 14 days of enrollment and are receiving high-flow nasal cannula, non-invasive ventilation, invasive mechanical ventilation, or ECMO at enrollment, in whom the current respiratory failure is thought to be due to SARS-CoV-2 infection and in whom respiratory support was initiated within 4 days prior to randomization.

Additional eligibility criteria apply for aviptadil and remdesivir. Participants who have received remdesivir prior to study entry are excluded from randomization to remdesivir.

The **primary endpoint** is a 6-category ordinal outcome that assesses the recovery status of the patient at Day 90, described in Appendix A, and is referred to as "**recovery"**. The categories of the ordinal outcome, from best to worst, start with 3 categories of "recovery" defined by the number of days alive at home and not on new supplemental oxygen, followed by 3 categories for "not recovered" defined as a) discharged but not to home or at home but still requiring continued new supplemental oxygen**,** b) hospitalized or receiving hospice care, and c) death at day 90.

Primary Objectives
1. To determine whether aviptadil is superior to placebo when given with standard of care for the primary outcome of recovery based on a 6-category ordinal outcome evaluated at 90 days after randomization.
2. To determine whether remdesivir is superior to placebo when given with standard of care for the primary outcome of recovery based on a 6-category ordinal outcome evaluated at 90 days after randomization.

Duration
The primary and most secondary outcomes will be collected during the first 90 days of follow-up. In addition, participants will be followed through 180 days for hospitalizations and deaths. SAEs that are related to study interventions will also be reported through 180 days.

Sample size
This Phase III trial is planned to provide 80% power to detect an odds ratio of 1.5 for improvement in recovery status at Day 90 for an investigational agent versus placebo, comparing treatment groups by intention to treat using a proportional odds model for the ordinal outcome. The planned sample size is 640 participants (320 per group) for each investigational agent versus placebo comparison. The sample size is not adjusted for inflation of Type I error due to multiple comparisons (separate tests for the two investigational agents).

The sample size may be re-estimated before enrollment is complete. The re-estimation will be performed by study personnel who are blinded to any data by treatment group.

Randomization Stratification
Randomization will be stratified by study design stratum (Figure 1), by disease severity (2 strata, defined by receipt of *mechanical ventilation or ECMO* at enrollment), and by study site pharmacy.

Monitoring
An independent DSMB will review interim data on a regular basis for safety and efficacy. Initially, monthly full reviews are planned, and weekly safety reviews for the aviptadil versus

placebo comparison. Prior to expanding enrollment to all sites, a full review by the DSMB will be conducted after approximately 40 participants have been randomized by the vanguard sites and have Day 5 data available.

After enrollment increases with the inclusion of non-vanguard sites, the DSMB will use asymmetric Haybittle-Peto monitoring boundaries for ***mortality*** to assess interim data for harm or benefit due to the investigational agents; the DSMB may recommend discontinuation of an investigational agent if the risks are judged to outweigh the benefits. No formal futility assessments are planned.

For an investigational agent, if the trial is stopped early, further enrollment of the investigational agent will be terminated if applicable, and the trial data for the investigational agent will be unblinded and reported with data through 90 days of follow-up. Follow-up of all participants will continue through 6 months using the data collection plan described in the master protocol.

1.3 Randomization

The randomization is described in section 6.1 of the protocol.

For the first two investigational agents, aviptadil and remdesivir, the study population consists of four "trial design strata", described in section 1.2 (Figure 1).

- Participants in stratum 1 are eligible for aviptadil and remdesivir, and have not used remdesivir prior to study entry. These participants will be randomized in equal proportions to one of the four treatment combinations in the 2x2 factorial formed by aviptadil/aviptadil placebo and remdesivir/remdesivir placebo.
- Participants in each of the other 3 trial design strata will be randomized 1:1 to the investigational agent versus matched placebo.

Within each design stratum, randomization will be stratified by study site pharmacy (several clinical sites may share one study site pharmacy) and by disease severity (receipt of *invasive mechanical ventilation or ECMO*) at entry. Within each randomization stratum, mass-weighted urn randomization[1] will be used to generate the treatment assignments, with equal allocations across the treatment groups. The number of treatment groups depends on the design stratum.

With this approach, participants will be equally allocated to aviptadil versus matched placebo (strata 1, 2, and 4) and to remdesivir versus matched placebo (strata 1 and 3).

1.4 Sample Size Estimates

The sample size calculations are aimed at the pairwise comparisons between each of the two investigational agents and its matched placebo. A total sample size of 800 participants is estimated.

To address the first primary objective, participants who are randomized to aviptadil will be compared to those randomized to the matched aviptadil placebo, pooled over design strata 1,

2 and 4. For the second primary objective, participants who are randomized to remdesivir will be compared to those randomized to the matched remdesivir placebo, pooled over design strata 1 and 3. Treatment groups will be compared by intention-to-treat for the 6-category ordinal outcome of recovery on Day 90 (primary outcome) using proportional odds models.

The planned sample size for each pairwise comparison is 640 participants (320 participants in each group, pooled across the corresponding design strata). The sample size is sufficient to detect an odds ratio (OR) of 1.5 with 80% power, using a two-sided test with a significance level of 0.05.

Participants in design stratum 1 (2x2 factorial) will contribute to both comparisons; assuming that 60%, 10%, 10%, and 20% will be enrolled in trial design strata 1-4, respectively, 800 participants in total will result in 640 participants for each of the two pairwise comparisons. The total sample size, which depends on the percentage of participants enrolled in the factorial design, will be periodically assessed by the protocol team.

Sample size calculations are described in detail in Section 6.3 of the protocol.

Blinded sample size re-estimation will be carried out before enrollment is complete to determine whether the planned sample size of 640 participants for each of the pairwise comparisons, followed for 90 days, will be sufficient to adequately power the trial. The blinded sample size re-estimation does not involve unblinding of the treatment difference. It will be based on the pooled outcome data, the relative enrollment into the two disease strata and four design strata, the number of withdrawals prior to the infusion, adherence to the blinded infusions, and the amount of missing data.

2 Interim DSMB Reviews: Goals and Format

Each investigational agent versus control will be reviewed as a separate clinical trial. For full DSMB reviews, a joint open report will be provided for aviptadil and remdesivir, with separate data summaries for the aviptadil and remdesivir cohorts (defined below), and key summaries by the four design strata. Separate closed reports will be provided for aviptadil and remdesivir, with similar layouts of the data summaries.

Analysis cohorts: Safety and efficacy analyses (closed reports) will be conducted for the Aviptadil and Remdesivir cohorts, and the Factorial cohort will be used to assess whether the effects of aviptadil and remdesivir are independent:

1. ***Aviptadil cohort:*** Participants who were randomized to aviptadil or its placebo (design strata 1, 2, and 4)
2. ***Remdesivir cohort:*** Participants who were randomized to remdesivir or its placebo (design strata 1 and 3)
3. ***Factorial cohort:*** Participants who were randomized to the four treatment combinations formed by the aviptadil/placebo and remdesivir/placebo pairs (2x2 factorial design; design stratum 1).

   For the factorial cohort, analyses will focus on assessing whether the effects of aviptadil and remdesivir are independent of each other, described in section 9.

Goals of the interim reviews:

- Protect the safety of study participants.
- Advise on stopping or modifying the trial for patient safety in case of emerging data on harm, or for efficacy in case of evidence of overwhelming benefit.
- After the first 40 participants are enrolled for the aviptadil vs placebo comparison, review safety and study implementation
- Review the conduct of the trial

Timing of the reviews: The DSMB will conduct frequent safety reviews. The initial safety review for either investigational agent will be conducted after approximately 40 participants are enrolled and have Day 5 data available, or earlier. Subsequent full reviews will be timed according to the recommendations of the DSMB and study leadership.

After the initial safety review, weekly safety reports for aviptadil versus placebo will be provided to the DSMB. At the discretion of the DSMB, the frequency of these (initially weekly) safety reports may be modified. ***The DSMB may request interim reports that are focused on safety at any time***.

Review meetings will typically consist of an Executive session (optional; closed), open session, closed session, and a second open session to give feedback to study leadership (optional).

Masking of treatment group labels in interim reports: In the open reports, any data reports will be pooled across all treatment groups. In the closed reports, treatment group labels will masked; for example as “Group A” through “Group D”. The treatment group labels will be consistent across all analyses and over subsequent reports. The DSMB will be unmasked to the treatment group labels.

Open report to the DSMB

The open report will contain:

- A synopsis of the trial design and current status of the platform trial
- Responses of the study team to DSMB requests
- A summary prepared by the study leadership
- Data summaries for enrollment
- Separate data summaries for the Aviptadil and Remdesivir cohorts:
    - Enrollment and baseline characteristics
    - Summary of adherence to infusions
    - Eligibility violations and protocol deviations
    - Summary reports for data completeness and study conduct
- Emerging external data, e.g., results of phase I or II trials on the investigational agent, will also be provided to the DSMB by the study leadership. This is usually included with the open report, but may be shared confidentially if needed.

All data summaries in the open report will be pooled across the treatment groups. The open reports will be prepared by the blinded statisticians in cooperation with the unblinded statisticians. In addition to the DSMB, open reports will be provided to the study team, and posted on the website for access by study investigators.

While the study is ongoing, summaries by treatment group, and comparisons of the investigational agents versus their placebo are restricted to the confidential closed report to the DSMB. Additionally, all summaries of follow-up data other than the data completeness and study conduct reports (pooled across treatment groups) will be restricted to the confidential closed report. For the **planned sample size re-estimations prior to completion**, pooled outcome data will be provided to the blinded study statisticians and study leadership. On a case-by-case basis, other pooled follow-up data may be provided if explicitly approved by the DSMB. ***Data that allow estimation of the treatment differences will remain blinded.***

Closed reports to the DSMB (full review)

A separate closed report will be provided for each investigational agent. All data summaries in the closed reports will be by (masked) treatment group. Closed reports for a full review will contain:

- Specific data summaries requested by the DSMB or study leadership
- Data summaries in the open report, by treatment group (enrollment, baseline characteristics, eligibility violations), described in sections 4 and 5.
- Data summaries to assess safety of the investigational treatment, described in sections 6 and 7. Data summaries for selected "efficacy outcomes" will also be included in each report, because these data contain information about the risk/benefit profile of the investigational agent. Analyses are described in section 8.
- Data summaries on data completeness and study conduct, described in section 11
- Interim monitoring boundaries for efficacy or harm (section 10)
- Listings of incident grade 3 and 4 adverse events, serious adverse events (SAE), protocol-specified exempt events (PSESE) described in Appendix C, unanticipated problems (UP), suspected unexpected serious adverse reactions (SUSAR), and deaths.
- Listings of early discontinuation of aviptadil or remdesivir (or matched placebo) with reason of discontinuation.

Closed Weekly Safety Report

Weekly DSMB reviews of safety data for the aviptadil versus placebo comparison will include the following data summaries:

- Summaries of the composite primary safety outcome of grade 3 or 4 AEs, SAEs, PSESEs, or death through Days 5 and 28, and its components
- Safety summaries for infusion reactions: infusion dose, peri-infusion grade 1-4 AEs, modifications in the infusion rates due to AEs, peri-infusion hypotension incidence, vasopressor use.
- Event listings for incident grade 3 and 4 AEs, SAEs, PSESEs, SUSARs, UPs and deaths (events that were reported since the previous review will be highlighted).
- Narratives for selected SAEs, SUSARs or UPs, particularly those judged related to study treatment.
- Incidence of grade 3 or 4 laboratory abnormalities

At the discretion of the DSMB, the frequency and content of these (initially weekly) safety reports may be modified.

3 Analysis Principles

Each investigational agent versus control will be treated as a separate clinical trial. Separate closed reports will be provided for each investigational agent and its corresponding randomized control group. The trial design does not allow to compare investigational agents (aviptadil and remdesivir) against each other. The pairwise comparisons of each agent versus control will **not** be adjusted for potential inflation of Type I error due to multiple comparisons.

The following principles apply for the comparisons of each investigational treatment against its randomized control arm (matched placebo).

Analysis populations for safety and efficacy outcomes:

- Comparisons for ***safety outcomes*** will be ***by modified intention-to-treat (mITT)***. The modified intention-to-treat analysis is restricted to participants who received a complete or partial infusion of the investigational agent/placebo; participants who did not receive ***any*** of the investigational agent/placebo are excluded.
- Comparisons for ***efficacy endpoints*** will be ***by intention-to-treat (ITT),*** unless otherwise stated. Sensitivity analyses by modified intention-to-treat will be carried out for primary outcomes and key secondary outcomes.
- Under certain circumstances, the efficacy analyses may be performed by mITT instead of ITT. For example, if enrollment is stopped due to a safety concern, it is customary to not initiate study treatment in participants who were randomized but did not yet start treatment. In this case, it would be appropriate to exclude such participants from efficacy analyses, because the reason for not starting treatment is independent of the treatment assignment.
  In general, prior to the unblinding of data, the blinded statisticians and study leadership will decide whether the efficacy analyses should be by mITT.

Analysis cohorts for individual investigational agents:

- ***Aviptadil cohort:*** The study population for the aviptadil versus placebo comparisons consists of design strata 1, 2 and 4; in stratum 1, participants are pooled across the two remdesivir arms.
- ***Remdesivir cohort:*** The study population for the remdesivir versus placebo comparisons consists of design strata 1 and 3; in stratum 1, participants are pooled across the two aviptadil arms.
- ***Factorial cohort:*** Participants who were randomized to the four treatment combinations formed by the aviptadil/placebo and remdesivir/placebo pairs (2x2 factorial design; design stratum 1).
  - The primary analysis to assess whether the effects of aviptadil and remdesivir are **independent** of each other will be conducted in the factorial cohort, by testing for an interaction effect. If there is evidence for an interaction ($p<0.05$), the effect of aviptadil versus placebo will be estimated for those who were randomized to remdesivir, and for those who were randomized to remdesivir-matched placebo. Analyses that will be conducted in the factorial cohort are described in Section 9.

Comment: Additionally, the independence of the effects of aviptadil and remdesivir will also be assessed in subgroup analyses within the aviptadil and remdesivir cohorts. For example, in subgroup analyses for aviptadil versus placebo, heterogeneity of the treatment effects across subgroups by remdesivir use (use at baseline or randomized to remdesivir versus neither) will be assessed by testing for an interaction between treatment and subgroup indicators. The assessment of independence is protected by randomization in the factorial cohort, but not in the subgroup analyses in the aviptadil or remdesivir cohorts.

Descriptive statistics will be reported overall and by randomized group. For categorical outcomes, the number and percent in each category will be reported; percentages will be of non-missing values, if data are not complete. Continuous variables will be summarized by median (interquartile range [IQR]) and/or mean (SD). Continuous variables may be categorized (e.g., age may be broken into categories to investigate the distribution across age groups).

Stratification: Tests comparing the investigational agent versus control for primary outcomes and key secondary outcomes will be stratified by disease severity (2 strata, by receipt of *invasive mechanical ventilation or ECMO* versus neither).

There are several exceptions:

- Early in the trial, analyses will be unstratified, until sufficiently many participants are enrolled such that each of the two disease severity strata contains at least 20 participants. This guideline on the minimal size of the strata aims to avoid unstable analyses and/or loss of power that may result from the use of sparse strata. In the case of time-to-event data, analyses will be stratified when sufficiently many events have accrued, e.g., at least 10 events per stratum.
- Sensitivity analyses exploring the effect of stratification will be provided for key analyses that may prompt the DSMB to recommend stopping or modifying the trial. In particular, extensive sensitivity analyses will be provided for the treatment difference in *mortality*, when test statistics approach the interim monitoring boundaries. Such sensitivity analyses will include the following:
    - Unstratified analyses
    - Stratification by study design stratum (3 strata in the aviptadil cohort, 2 strata in the remdesivir cohort)
    - Stratification by disease severity and study design stratum (6 strata in the aviptadil cohort, or 4 in the remdesivir cohort)
    - Additional stratification by geographical region (U.S, Europe, other), provided individual strata are sufficiently large (20 participants or more).

For time-to-event analyses, stratification by disease severity usually implies separate baseline hazard functions for the two strata; if event numbers in strata are too small, however, the strata indicator may be included in the model as an additive covariate instead.

Comment: Randomization in TESICO will be stratified by disease severity, by study site pharmacy, and by the four design strata. The randomization strata were designed to ensure balanced treatment groups. Statistical analyses (treatment comparisons) will be stratified by disease severity only (unless specified otherwise), in order to avoid potentially unstable analyses due to small strata.

For **binary outcomes**, probabilities will be compared between the investigational agent and its control group using Cochran-Mantel-Haenszel tests (CMH) or logistic regression. The CMH tests will be stratified by disease severity, as described above under "stratification". Odds ratios (OR) with 2-sided 95% confidence intervals (CI) will be estimated using logistic regression models.

For longitudinally measured binary outcomes, the treatment effect through follow-up will be estimated with 95% confidence intervals using generalized estimating equations (GEE) with a logit link function; the treatment effect is estimated via the interaction between the indicator for treatment group and the indicator for follow-up (versus baseline) visits. When there is more than one follow-up visit, "visit number" (day) may be included as categorical variable in the model, for variance reduction; alternatively, "time" may be included as a continuous variable.

Ordered categorical outcomes (e.g., recovery) will be compared between treatment groups using proportional odds models, and the summary OR will be estimated with a 2-sided 95% CI.[2] Additionally, to aid the interpretation, the ordinal outcome will be dichotomized according to cumulative probabilities of the ordered categories, comparing treatment groups for proportions of participants in category 1, in the "best 2 categories", "best 3 categories", etc.; these comparisons will be performed using logistic regression (or stratified CMH tests).

Models will be adjusted for disease severity at study entry (2 categories, by use of *invasive mechanical ventilation or ECMO* versus neither), by including the corresponding indicator variable in the model.

The validity of the proportional odds assumption will be assessed by testing for heterogeneity in the log ORs (for the treatment effect) across the dichotomized cumulative ordered categories in the corresponding logistic regression model (partial proportional odds model, test for "unequal slopes").

- The primary sensitivity analysis testing the proportional odds assumption will compare the unadjusted proportional odds model for the treatment comparison (null model) versus a partial proportional odds model that allows for "unequal slopes" across the dichotomized cumulative categories (i.e., when testing the proportional odds assumption for the treatment comparison with respect to the recovery outcome on a given day, the model will allow for heterogeneous ORs across the outcome categories) as well as across the stratification covariate (i.e., the strata defined by disease severity) (full partial proportional odds model).

Continuous outcomes will be compared between treatment groups using ANCOVA models for comparing means, if the ANCOVA model assumptions hold. If the distributions of the continuous outcomes are skewed, outcomes may be transformed, or compared between treatment groups using rank-based methods, such as the Wilcoxon test, or quantile (median) regression. For example, biomarker levels often require log-transformation to meet model assumptions for ANCOVA analyses.

Comparisons between treatment groups for a continuous outcome will be adjusted for baseline values of the outcome, for the purpose of variance reduction, unless there are

concerns over model stability with such an adjustment. For this purpose, the baseline value will be included as covariate in the model (e.g., ANCOVA, linear mixed models).

To estimate the treatment effect for *longitudinally measured continuous outcomes*, the outcome will usually be defined as "change from baseline" (difference at follow-up visit minus baseline value). The treatment effect through follow-up will then be estimated with 95% confidence intervals using generalized estimating equations (GEE) with an indicator for treatment group, or, in the case of Gaussian responses, the corresponding mixed effects models with random effects for participants. To improve the model fit and reduce error variance, "visit number" (day) may be included as categorical variable in the model; alternatively, "time" may be included as continuous variable. Models will also be adjusted for the baseline values of the outcome variable.

Time-to-event outcomes will be summarized with Kaplan-Meier estimates for cumulative probabilities over time, and compared between treatment groups using stratified log-rank tests or Cox proportional hazards models, or their competing risk analogues.

In case "death" is a competing risk for the outcome (e.g., for time to hospital discharge), the following competing risk methods will be used:

- Aalen-Johannsen estimator for the cumulative incidence function (analogue to the Kaplan-Meier estimate)[3]
- Gray's test with $\rho=0$ (analogue to the log-rank test)[4]
- Fine-Gray estimates and tests for the sub-distribution hazard ratio (analogue to the Cox proportional hazards model).[5,6]

The proportional hazards assumption will be tested by adding an interaction term for time by treatment group to the model. The cumulative proportions of participants who experienced the event will also be compared at given time points (specified in secondary objectives, e.g., at 28 days); in this case, the cumulative proportions will be estimated using Kaplan-Meier estimates or the competing risks analogue, and/or as proportion of participants who reached the time point (e.g., time since randomization $\geq$ 28 days).

The **administrative follow-up time** is defined as the minimum of (cut date minus randomization date) or the analysis time period. For example, the analysis time period for the time to hospital discharge is 180 days, and the analysis time period for the important safety endpoint, the composite of *grade 3 and 4 events, SAEs, PSESEs, or death*, is 5 days or 28 days. The **administrative censoring date** is the earlier of the cut-date of the dataset or the randomization date plus analysis time period.

> **Comment:** The notion of "administrative censoring" is important in time-to-event analyses in the presence of competing risks. For example, the Fine-Gray method for estimating the sub-hazard ratio for time to hospital discharge can be approximated by using a Cox proportional hazards model where follow-up time for participants who died without being discharged is not censored at death, but is carried forward to the administrative censoring date with event status "not recovered".

Censoring for time-to-event analyses

For **interim** analyses, the type of censoring used will depend on the data collection schedule.

- If the reporting of the endpoint is data-driven (e.g., SAEs and deaths are reported as they occur), then follow-up is censored at the administrative censoring date, at the date of withdrawal, or loss to follow-up, whichever occurs earliest.
- If the date of the event is elicited retrospectively at fixed study visits spaced more than one week apart, follow-up will be censored at the last day the endpoint status was ascertained. For example, this applies to endpoints that require information on whether the patient has been "at home" for a given period of time, such as time to *sustained recovery*.
- Sensitivity analyses will be provided for key analyses when the outcome status is uncertain.

For **final** analyses, follow-up will be censored on the last day the outcome status was ascertained.

Adverse events (AEs) will be classified by system organ class according to MedDRA®[1] (currently version 24.0 [March 2021] is used; when new versions are implemented, items are recoded). AEs will be graded according to the *DAIDS Table for Grading the Severity of Adult and Pediatric Adverse Events*, *Corrected Version 2.1 (July 2017)* (also referred to as the *DAIDS AE Grading Table*).[7] Hypotension AEs will be graded according to Table D-3 in Appendix D; this table is replicated from the TESICO protocol, version 2.0, Table 5. Cause of death will also be coded according to MedDRA®.

> **Comment:** Under version 1 of the TESICO protocol, hypotension AEs were graded according to the DAIDS AE Grading Table.

The number and percent of participants with peri-infusion grade 1-4 AEs will be summarized by day and grade, and by type and grade. The percentage of participants with AEs will be compared between treatment groups according to grade cut-offs, e.g., "percent of participants with any AE", "percent of participants with grade 2 or higher AEs", etc., using CMH tests. The total number of events and median (IQR) number of events per participant will also be summarized.

Additionally, the incidence of grade 3 and higher AEs will be summarized (number and percent of participants) by MedDRA® System Organ Class and grade using stratified CMH tests. Incidence of grade 3 and 4 AEs will also be compared between treatment groups as part of composite safety endpoints using time-to-event methods.

Significance level, two-sided tests: Unless noted otherwise, statistical tests and confidence intervals will be 2-sided, confidence intervals will have approximate 95% coverage probability, and test results with P-values $\leq$ 0.05 will be considered "significant".

Cut-date for interim reviews: Analysis data sets will be locked several days (or weeks) prior to the review date, to allow the unblinded statisticians time to prepare a consistent report.

[1] The Medical Dictionary for Regulatory Activities terminology is the international medical terminology developed under the auspices of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). MedDRA® is a registered trademark of the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA)

The cut-date may be earlier than the date of the data lock, to allow for lag time in the reporting of events. Early in the trial, the cut date and lock date will be very close to the review date, to ensure timely safety reviews.

4 Enrollment and Eligibility

For the open report, the following enrollment and eligibility summaries will be provided:

- Enrollment over calendar time: plot by day or week, cumulative and increments.
- Enrollment by study design stratum (4 strata), and pooled across the relevant design strata for the aviptadil/placebo and remdesivir/placebo comparisons.
- Enrollment by site pharmacy and by country: number (%)
- Eligibility: number (%) and reasons for eligibility violations

These summaries will be provided overall, and by disease severity randomization stratum.

For the closed report, enrollment and eligibility violations will be summarized by treatment group.

5 Baseline Characteristics

Baseline characteristics will be based on information collected on baseline and screening forms. Separate summaries will be provided for the Aviptadil and Remdesivir analysis cohorts.

For the open report, data will be pooled across treatment groups. In addition to the overall summaries, selected baseline characteristics will also be summarized by disease severity at study entry (2 strata, by use of *invasive mechanical ventilation or ECMO*).

For the closed report, baseline characteristics will be summarized by treatment group. For interim closed reports, baseline characteristics will usually be summarized by **modified intention-to-treat (mITT),** for consistency with the safety summaries.

Unless noted otherwise, categorical variables will be summarized with the number and percentage (N, %) of participants in each category, and continuous variables will be summarized with the median and interquartile range (IQR); in the open report, in addition, the mean (SD) and range may be provided.

The following baseline characteristics will be reported:

- Number of participants with baseline data, overall and by study design stratum (N, %)
- Demographics
  - Age: distribution in categories 18-39, 40-49, 50-59, 60-69, 70-79, ≥80 years; and summary as continuous variable
  - Sex at birth
  - Ethnic group: Asian, Black, Latino/Hispanic, White, other
  - Country of enrollment

- Type of residence prior to COVID-19 ("home")

- COVID-19 related characteristics
  - Duration of symptoms prior to enrollment
  - Level of care: non-ICU versus ICU
  - Respiratory support (4 categories: high-flow nasal cannula [HFNC], non-invasive ventilation [NIV], invasive mechanical ventilation without ECMO, or ECMO)
  - Time since initiation of oxygen support
  - Extent of lung infiltrates (3 categories: none, unilateral, bilateral)
  - ARDS (defined as bilateral lung infiltrates and SF ratio <315)
  - Receipt of SARS-CoV-2 vaccination (N, %)
    - Fully vaccinated at the onset of COVID-19 symptoms (≥ 14 days after the final vaccine dose [≥ 14 days since second dose, or since first dose if only one dose is required])
    - Type of vaccine
    - Received as part of a blinded clinical trial
- Vital signs (median [IQR] and categories)
  - Respiratory rate (categories: ≤20 vs >20 breaths/minute)
  - Oxygen saturation (SpO2) (categories: <92%, 92-96%, >96%)
  - Fraction of inspired oxygen (FiO2) (categories: <0.30, 0.31-0.40, 0.41-0.70, >0.70)
  - SpO2:FiO2 (SF ratio) (categories: <315, ≥315)
  - Temperature (categories: <38º C, ≥ 38º C)
  - Heart rate (categories: <100, ≥100 bpm)
  - Systolic blood pressure (SBP) (categories: <90 mmHg, 90-110 mmHg, >110 mmHg)
  - Mean arterial pressure (MAP) (4 categories: <65 mmHg with vasopressor use, <65 mmHg without vasopressor use, ≥65 mmHg with vasopressor use, ≥65 mmHg without vasopressor use)
  - Current vasopressor use
  - ARDS: SF ratio < 315 and bilateral pulmonary infiltrates
- Acute organ dysfunction during the index COVID-19 illness
  - Cardiac and vascular dysfunction
  - Hematologic dysfunction
  - Hepatic decompensation (exclusion criterion for aviptadil)
  - Serious infection other than SARS-CoV-2 (respiratory and non-respiratory)
  - Neurologic dysfunction
  - Renal dysfunction
- History of chronic conditions (present prior to the index COVID-19 illness)
  - Compromised immune function, defined as current use of antirejection medication after transplant, cytotoxic chemotherapy, or treatment with biological medicine for autoimmune disease or cancer; HIV; or immunosuppressive disorder other than HIV.
  - Metabolic/vascular co-morbidities, defined as history of diabetes mellitus requiring treatment, a cerebrovascular event (thrombotic or hemorrhagic), heart failure, or an MI or other acute coronary syndrome, overall and by components
  - Hypertension with and without history of other metabolic and vascular co-morbidity (4 groups):
    i) no hypertension or other metabolic/vascular co-morbidity;
    ii) hypertension without metabolic/vascular co-morbidity;
    iii) metabolic/vascular condition without hypertension; and
    iv) hypertension and a metabolic/vascular co-morbidity.

Metabolic/vascular co-morbidities include a history of diabetes, a cerebrovascular event, heart failure, or an MI or other acute coronary syndrome
    - Renal impairment or requirement of renal replacement therapy, overall and by components
    - Chronic obstructive pulmonary disease (COPD)
    - Chronic continuous supplemental oxygen use
    - Hepatic impairment
    - Cancer
- Other clinical characteristics
    - BMI (<30, 30-39.9, 40+)
    - Pregnancy, and gestational age (remdesivir cohort only, as pregnancy is an exclusion criterion for aviptadil)
- COVID-19 treatments
    - Receipt of remdesivir prior to randomization (N, %), and number of days
    - Corticosteroid use, summarized overall, and by disease severity at study entry (receipt of *invasive mechanical ventilation or ECMO*)
    - Antiplatelet/anticoagulant therapy (none; prophylactic heparin; intermediate or therapeutic heparin or other anticoagulant therapy; aspirin; or other antiplatelet therapy)
    - Immune modulators (IL-1 inhibitors, IL-6 inhibitors, interferons, JAK inhibitors, TNF inhibitors, other)
    - Convalescent plasma for SARS-CoV-2 infection
    - Hyperimmune intravenous immunoglobulin (hIVIG)
    - Neutralizing monoclonal antibodies for SARS-CoV-2
- Pulmonary vasodilators, by type
- Sedatives, by type
- Blood pressure lowering medications, by type
- Other concomitant medications
    - Antidiarrheals
    - Antifungals
    - Antirejection medications
    - Biologics for cancer or autoimmune disease
    - Cytotoxic chemotherapy
    - NSAIDs (at least 7 days)
- Laboratory values: as continuous outcomes, and number (%) of grade 3 or 4 lab abnormalities according to the *DAIDS AE Grading Table.*
- Co-enrollment in other trials, by trial
- Genomics consent

Some biomarkers will be measured centrally from stored samples, for example, SARS-CoV-2 antigen and antibody levels in plasma and SARS-CoV-2 viral RNA from nasal mid-turbinate swabs. If these measures are available, they will be included in interim reports.

Open report only:
In addition to the overall summaries, selected baseline characteristics will also be summarized by study design stratum, including age, gender, race/ethnicity, geographic region, disease severity, admission to ICU, duration of symptoms prior to enrollment, and duration of support for respiratory failure.

6 Administration of Study Treatment

These data are an important part of the safety review, with particular emphasis on infusion-related reactions and symptoms occurring during or within up to 2 hours after the infusion. These reactions and symptoms will be graded according to the DAIDS AE Grading Table.

The administration of study treatment is also an essential element of study conduct. The data summaries described in this section will be provided:

- In the closed report to the DSMB, by treatment group
- In the open report, selected summaries describing adherence, pooled across treatment groups

Summaries of AEs or infusion-related reactions are restricted to the closed report. Selected summaries will also be provided separately for the two disease severity strata.

Analyses will be by modified intention-to-treat (mITT), unless specified otherwise. The treatment comparisons will be performed using the methods described in section 3 for binary and continuous outcomes: stratified CMH test or logistic regression for comparing percentages, Wilcoxon rank-sum test [or quantile regression] for comparing medians, ANCOVA models for comparing means.

6.1 Infusion of Aviptadil (Active or Placebo)

Study population *(Aviptadil cohort, mITT)*: Unless otherwise noted, the study population consists of participants who were randomized to aviptadil or its placebo, excluding those who did not receive **any** aviptadil/placebo (modified intention-to-treat [mITT]).

Aviptadil (and its placebo) will be administered by intravenous infusion for 3 days, over a continuous 12 hour period each day. The protocol-specified infusion rate for the first infusion day is 50 pmol/kg/hr (Day 0), to be increased to 100 pmol/kg/hr on Day 1 and 150 pmol/kg/hr on Day 2. This corresponds to a protocol-specified dose of 600, 1200, or 1800 pmol/kg for infusions on Days 0-2. Infusions may be paused or discontinued due to side effects, or the target dose may be decreased.

In the following, "aviptadil" refers to the blinded aviptadil infusion (active or placebo). "Peri-infusion" refers to the time period during or within 2 hours after the infusion.

The following statistics will be used to summarize the infusions in each treatment group (active and placebo) for the closed reports, or pooled across treatment groups for open reports shared with investigators. For the closed report, treatment groups will be compared for the various outcomes using methods described in section 3.

For all participants who were randomized ***(Aviptadil ITT cohort)***:

- Non-administration of aviptadil:
  Number and percentage of participants who did not receive any aviptadil, and reasons

All other analyses will use the ***Aviptadil mITT*** cohort (i.e., analyses will exclude those who did not receive any blinded aviptadil in the study).

- For each of the Days 0-2: number and percentage of participants who did not receive any aviptadil, and reasons
- Adherence to the study treatment on Days 0-2 will be summarized over time: percent of participants who received no infusion, some (<50% of the day's protocol-specified infusion dose), most (50-90%) and full infusion (>90%) for each of the three days, and overall.
    - For each day, for each participant, the "percent dose infused" will be the estimated dose delivered expressed as percentage of the protocol-specified dose for the day (600, 1200, or 1800 pmol/kg).
    - For the summary across Days 0-2 ("overall"), categories are formed by the average of the three daily percent doses infused.
- Day of first infusion (same day as randomization, next day, or > 1 day after randomization), and time between randomization and start of infusion (median hours, IQR).
- For each of the Days 0-2: **Infusion of aviptadil**, rate, dose, and infusion time
    - Pre-infusion summary:
        - starting (patient-specific planned) infusion rate (categories: at versus below the protocol-specified rate for the day)
        - blood pressure (MAP, SBP)
        - use of vasopressors and dose
        - receipt of antidiarrheal agents to prevent infusion reactions
    - Dose infused (as percent of goal): median, 25th and 75th percentile, and according to categories (full [>90% of goal]; most [50-90%]; some [<50%]; no infusion)
    - For participants who received an incomplete infusion of the study drug on any day, reasons for the incomplete infusion (all that apply)
    - Timing of administration
        - Duration of infusion: median, 25th and 75th percentile of infusion time
        - Infusion start times (8am-12:00 noon; 12:01pm- 6:00pm, > 6pm)
    - Decrease in the infusion rate during the day (overall: worst category across Days 0-2; and worst category by day) (closed report only):
        - No decrease
        - Flow rate decreased, but no intermittent stop. The reasons for the decrease will be summarized (due to AE versus other reasons)
        - Infusion was paused (stopped but resumed), and not discontinued prematurely for the day; and reasons (AE versus other)
        - Premature discontinuation for the day (infusion duration <12 hours, and patient-specific planned dose not delivered), and reasons for discontinuation (AE or other)
        - Permanent discontinuation, and reasons
    - Listings of participants for whom the infusion was discontinued early, with reason for discontinuation, dose infused per day, disease severity at baseline, age, sex.
    - Concomitant medications peri-infusion (i.e., during or within 2 hours after the infusion) (overall: any use peri-infusion on Days 0-2; and by day)
        - Vasopressors, with dose in norepinephrine equivalent (NE)
        - IV bolus or colloids
        - Other medications to treat AEs, as collected on the infusion eCRF

Safety: infusion-related signs/symptoms, AEs, and lab markers (closed report only)
The primary safety outcome of *grade 3 or 4 AEs, SAEs, PSESEs, or death through Day 5* includes potentially infusion-related outcomes; the analysis of the primary safety outcome is described in section 7. The current section describes safety analyses that are focused on the peri-infusion time period (i.e., during and within 2 hours after the infusion).

Peri-infusion signs and symptoms are reported as AEs (with grade and action taken) on eCRFs. Local labs are reported at baseline, Days 1 and 2 for all participants, and Days 3 and 5 if clinically available. Labs on Day 5 are reported for participants who are treated in the ICU.

Side effects of aviptadil observed in previous studies include:

- Hypotension with and without vasopressor
- Diarrhea
- Facial flushing
- Bradycardia
- Tachycardia

For the aviptadil versus placebo comparisons, incidence and management of hypotension in particular will be monitored in detail. Blood pressure is recorded prior to the start of the infusion and every 2 hours peri-infusion on each infusion day. Vasopressor use and dose are reported: prior to the start of the infusion, the highest dose peri-infusion, and the dose at 2 hours after the infusion.

The following data summaries will be provided overall for Days 0-2, and separately by day. Treatment groups will be compared using statistical methods described under “analysis principles” in section 3; comparisons will be stratified by disease severity at study entry (2 categories, by use of *invasive mechanical ventilation or ECMO*).

- Number and percentage of participants with infusion-related signs or symptoms (reported during the infusion or within 2 hours after the infusion), by type and grade, by type and grade cut-off (i.e., $\geq$ grade 2, $\geq$ grade 3, etc.), and by type and action taken.

  Comments:
  1. For aviptadil, ***diarrhea is not included in the composite safety outcome*** of incident grade 3 or 4 AEs, SAEs, PSESEs or death unless it is a SAE or leads to discontinuation of the infusion (for the day, or permanent discontinuation). Therefore, peri-infusion diarrhea AEs will be summarized in two ways: (1) as reported, and (2) restricted to diarrhea AEs that lead to discontinuation of the investigational agent.
  2. Hypotension, diarrhea, facial flushing, bradycardia, and tachycardia are collected as part of the infusion-related signs and symptoms, and will be summarized by type and grade, and by type and action taken in response to the AE, as described above.

- **Peri-infusion hypotension summary**
  - Peri-infusion hypotension AEs reported on infusion eCRF (by grade: highest grade on Days 0-2, or highest grade on the day for the day-specific summaries)
  - Blood pressure (across days 0-2, and on each day)

- MAP decrease by > 20 mmHg from daily pre-infusion baseline
- Lowest peri-infusion MAP (mean [SD] across participants)
- Incidence of MAP < 65 mmHg (compared to daily pre-infusion baseline)
- Percentage of peri-infusion BP data points with MAP <65 mmHg (mean [SD] of percentage across participants)

o Infusion modifications for hypotension (prevalence on ≥ 1 day)
- Infusion not attempted for hypotension/vasopressor use
- For those with any infusion modification, summary of the highest-intensity infusion modification (across Days 0-2, and for each day): Rate decreased but infusion not paused; infusion paused but resumed; infusion discontinued for the day; infusion discontinued permanently

o Vasopressor use peri-infusion
- New peri-infusion vasopressor use on any day among patients not receiving vasopressors pre-infusion on Day 0 (Number in subgroup, N, %)
- New or increased peri-infusion vasopressor use on any day among patients who received vasopressors pre-infusion on Day 0 (Number in subgroup, N, %)
- Maximum peri-infusion vasopressor rate increase within a study day from pre-infusion to peak (in NE units; max over Days 0-2, then mean across patients; 0 NE is imputed for time points when no vasopressor is used. If the maximal peri-infusion vasopressor dose is lower than the pre-infusion dose, then 0 NE will be imputed).
- Percent with vasopressor rate higher at 2 hours post infusion than pre-infusion
- Incidence of vasopressor rate increase by > 0.03 mcg/kg/min NE units peri-infusion relative to daily pre-infusion baseline (N, %)
- Incidence of peak (absolute) vasopressor rate of > 0.1 mcg/kg/min NE units

o IV fluid use peri-infusion in response to hypotension AE
- IV fluid (crystalloid ≥ 500 mL or equivalent colloid volume) peri-infusion, prevalence
- For those who received IV fluid or colloid: maximum IV fluid volume on one day (max across days for each participant, then mean across participants)

- Summary of *SBP and MAP trajectories*, from infusion start to 2 hours after infusion
  - Mean trajectories will be plotted by treatment group, and compared using longitudinal models.
  - For selected participants, *individual trajectories* for MAP and SBP will be plotted for each of the infusion days. In particular, individual trajectories will be plotted for participants who used vasopressors, who experienced peri-infusion hypotension (MAP < 65 mmHG or hypotension AE), or for whom the infusion flow rate was decreased, the infusion paused, or discontinued. These events will be marked on the individual trajectories. Thus, the individual trajectories serve as “line listings” for hypotension events.

- Heart rate:
    - Mean trajectories over time, similar to MAP
    - Percent of participants for whom the peri-infusion heart rate decreased to below 60 bpm at any time, stayed between 60-100 bpm at all times, or reached >100 bpm at any time.

- Medications (other than vasopressors) received in response to AEs during or within 2 hours after infusion, number and percentage of participants and type of medication

- Listings of participants who died or experienced grade 3 or 4 AEs, SAEs, or PSESEs are provided as part of the safety analyses described in section 7.

6.2 Infusion of Remdesivir (Active or Placebo)

Study population *(Remdesivir cohort, mITT)*: Unless otherwise noted, the study population consists of participants who were randomized to remdesivir or its placebo, excluding those who did not receive **any** remdesivir/placebo (modified intention-to-treat [mITT]).

Remdesivir (and its placebo) will be administered once-daily by intravenous infusion (over 30 minutes) for up to 10 days, or until hospital discharge, whichever comes sooner. Remdesivir will be administered on Day 0 as a 200 mg IV loading dose, followed by a 100 mg maintenance dose on subsequent days. Remdesivir may be discontinued after 5 or more days, per discretion of the treating clinician, if the participant is no longer requiring respiratory support.

In this section, "remdesivir" refers to the blinded remdesivir infusion (active or placebo).

The following statistics will be used to summarize the infusions in each treatment group (active and placebo) for the closed reports, or pooled across treatment groups for reports shared with investigators:

- Number and percent of participants who received (any) remdesivir/placebo, by day
- Number of days remdesivir was administered: median, IQR, distribution.
- Number and percent of participants for whom remdesivir was discontinued, and reasons for discontinuation.
- Number and percent of participants for whom the complete target volume was not administered, by day; and reasons for incomplete administration (pooled over days).
- Number and percent of participants for whom a daily remdesivir infusion was discontinued prematurely, by day, and pooled across days.
- Pooled across days:
    - Number and percentage of participants with infusion-related signs or symptoms (reported during the infusion or within 2 hours after the infusion), by type and grade, and by type and action taken. (Closed report only)
    - Number and percent of participants for whom a remdesivir infusion was discontinued prematurely due to an AE

- Number and percent of participants for whom medications were prescribed during or within 2 hours following the infusion *in response to an AE*, and type of medication

7 Safety Analyses

The planned timing of safety reviews is described in section 2. An overview of the safety data collection is provided in Appendix D.

Analysis cohorts: Safety analyses will be conducted by modified intention-to-treat (mITT), unless otherwise stated, for the following three cohorts:

1. ***Aviptadil cohort, mITT:*** Participants who were randomized to aviptadil or its placebo, excluding those who did not receive any aviptadil/placebo
2. ***Remdesivir cohort, mITT:*** Participants who were randomized to remdesivir or its placebo, excluding those who did not receive any remdesivir/placebo
3. ***Factorial cohort, mITT:*** Participants who were randomized to the four treatment combinations formed by the aviptadil/placebo and remdesivir/placebo pairs (2x2 factorial design), excluding those who did not receive any of the aviptadil/placebo OR any of the remdesivir/placebo.
   In the factorial cohort, the presence of interactions between aviptadil and remdesivir will be assessed, for the primary safety endpoint and other key outcomes. If there is evidence for interactions ($p<0.05$) in a given outcome, the analyses described in Section 9 will be performed.

Comment: Because the safety profile of remdesivir has been well-described, safety summaries for the remdesivir cohort will be provided only after more than 40 participants are randomized to remdesivir vs placebo; the frequency of the safety reports for remdesivir will be determined by the DSMB.

A comprehensive safety review includes:

- Comparison of the treatment groups for the primary safety endpoint, its components, and analyses of secondary safety outcomes (described in this section)
- Analyses of infusion-related reactions and symptoms, described in section 6
- Evaluation of selected efficacy outcomes (e.g., recovery at Day 90, time to recovery, time to hospital discharge), which contain important safety information. Described in section 8.

In addition to the full DSMB reviews, more frequent, shorter safety reports will be provided to the DSMB, for example, weekly safety reports early in the trial.

This section describes the primary safety outcome, and the analyses of AEs, SAEs, UPs, SUSARs, and deaths. Comparisons between treatment groups will be stratified by disease severity at study entry (as described in section 3 under “stratification”).

In order to streamline the reporting of events, it was decided that certain protocol-specified exempt events (PSESE) are *not reported as SAEs*, unless they are considered related to the study treatment by the investigator. The PSESE in TESICO encompass a collection of serious events that are expected to occur commonly in the target population even in the absence of study interventions. While the PSESEs in this protocol are similar in severity to

SAEs, PSESEs are reported not on the SAE eCRF, but are reported as study endpoints on various other eCRFs. AEs that are considered PSESEs are listed in Appendix C. The composite outcome of *clinical organ failure or serious infections*, defined in Appendix B, is comprised of all PSESEs, except all-cause mortality (death is a PSESE, but is included in the composite of *clinical organ failure or serious infections* only if the cause of death corresponds to one of the components listed in Appendix B).

7.1 Safety Analyses for the Aviptadil and Remdesivir Cohorts

The following safety and tolerability outcomes will be analyzed; models will be stratified by disease severity (receipt of invasive mechanical ventilation or ECMO at study entry), as described in section 3 under "stratification", unless noted otherwise:

- The **primary safety endpoint** is a composite of incident grade 3 or 4 AEs, SAEs, PSESEs, or death through Day 5. The number and proportion of participants experiencing one of these events up through Day 5 will be tabulated, and treatment groups will be compared using a CMH test stratified by disease severity at study entry (2 categories, by receipt of *invasive mechanical ventilation or ECMO* vs. neither). The OR comparing the investigational treatment versus placebo will be estimated with a 95% CI using a logistic regression model that includes the treatment group indicator and the indicator for disease severity at study entry.
    - Mortality will be analyzed as a key secondary outcome, see below.
    - The individual components of the composite outcome will be summarized.
    - Sensitivity analyses for the primary safety outcome: After completion of enrollment, if the Day 5 status for the primary safety outcome is unknown for more than 2% of participants in the mITT cohort, then treatment groups will also be compared for time to event through Day 5 using a log-rank test, stratified by disease severity at study entry; the HR will be estimated with a 95% CI using a stratified Cox proportional hazards model, and the cumulative proportion of participants with events over the first 5 days in each treatment group will be estimated using Kaplan-Meier curves.

  Comments:
  1. For aviptadil, ***peri-infusion diarrhea is not included in the composite safety outcome*** of *incident grade 3 or 4 AEs, SAEs, PSESEs or death* unless it is a SAE or leads to discontinuation of the infusion (Appendix H1 of the TESICO protocol). After completion of the infusion, all incident grade 3 or 4 diarrhea AEs are included.
  2. Because most participants in the remdesivir cohort will also be randomized to aviptadil versus placebo, we will use the same composite safety outcome for the remdesivir versus placebo comparison as for aviptadil (i.e., diarrhea occurring during and up to 2 hours after the aviptadil/placebo infusion will be excluded as described above).

- All-cause mortality through follow-up will be analyzed using time-to-event methods. Cumulative proportions of participants who died in each treatment group will be estimated using Kaplan-Meier estimates, and summarized in tables (proportion of participants who died by Days 5, 14, 28, 60, 90, month 6) and figures (Kaplan-Meier curves with pointwise 95% CIs). Treatment groups will be compared for time to death using log-rank tests,

stratified by disease severity at study entry, and HRs will be estimated with 95% CIs using stratified Cox proportional hazards models.

- Cause of death will be MedDRA® coded and summarized by treatment group.

- The following composite endpoints will be analyzed using time-to-event methods (cumulative proportions of participants with events will be estimated using Kaplan-Meier curves with pointwise 95% CIs; treatment groups will be compared using log-rank tests; numbers and percent of participants with events will be summarized by treatment group, and overall HRs with 95% CI will be estimated using Cox proportional hazards models):
    - Composite of incident grade 3 or 4 clinical adverse events, SAEs, PSESEs, or death through Day 28
        - Components of the composite endpoint will be also be summarized.
        - In addition to time-to-event analyses, the treatment groups will be compared for the proportion of participants who experienced the composite endpoint by Day 28 using a stratified CMH test, similar to the primary safety analysis on Day 5.
    - Composite of SAEs, PSESEs, or death through Day 28 and Day 90
    - Composite of hospital re-admission or death through Day 90 and Month 6.

- Grade 3 or 4 AEs, SAEs, and UPs will be classified by MedDRA® system organ class. AEs will be graded for severity according to the *DAIDS AE Grading Table*, except for peri-infusion hypotension AEs, which are graded according to Table D-3 in Appendix D.

- Incident **grade 3 and 4 clinical AEs** are reported through Day 28. (A grade 3 or 4 AE is considered "incident" if the event was not present at baseline or increased to grade 3 or 4 from grades 1 or 2, or increased to grade 4 from grade 3.)
    - AEs that were reported to have occurred on Day 0 prior to the first infusion will be considered "baseline" and thus will be excluded from the analysis of incident AEs.
    - Grade 3 and 4 AEs that occur peri-infusion (i.e., during and within 2 hours after the infusion) on Days 0-2 and are reported on the infusion eCRFs will be included as incident grade 3 or 4 AEs, unless noted otherwise.
    - The number and percent of participants with incident grade 3 and 4 AEs will be summarized by MedDRA® system organ class and grade, and by MedDRA® system organ class and grade cut-off (i.e., grade $\geq$3, grade 4). Comparisons between treatment groups will be for the proportion of participants who experienced AEs of grade 3 or higher through Day 28, using stratified CMH tests or logistic regression, stratified by disease severity at study entry (*invasive mechanical ventilation or ECMO* vs. neither). Treatment groups will be compared for incidence of grade 3 or 4 AEs overall, and by system organ class.
    - System organ classes may be split up into MedDRA® preferred terms (PT) for the most frequent system organ classes, particularly for classes where the treatment difference is significant.
    - Other clinically meaningful AE groupings (beyond system organ class) may be developed by the study team, who are blinded to the treatment effect.

- Grade 1-4 clinical AEs are reported at baseline (Day 0 prior to infusion of the investigational agent), on Days 0-2 peri-infusion (during and within 2 hours after the

infusion), as well as on Days 14 and 28. The peri-infusion AEs are collected as "signs and symptoms" via checklist on the infusion eCRFs.

The analysis of peri-infusion AEs is described in section 6. In particular, the number and percent of participants with peri-infusion AEs will be summarized by day and grade, by type and grade, and by type and action taken (i.e., modification or discontinuation of the infusion).

- Treatment groups will be compared for the proportion of participants who developed **PSESEs** through Day 28 and through Day 90, using stratified CMH tests. In addition to the overall comparison, individual components of the composite PSESE outcome will be tabulated, and compared between treatment groups using stratified CMH tests. The components of the PSESE outcome are listed in Appendix C.
  Comment: The composite outcome of *clinical organ failure, serious infections, or death* is identical to the composite of all PSESEs; individual components of clinical organ failure are listed in Appendix B.

  Treatment groups will also be compared for the incidence of PSESEs using time-to-event methods; because death is a PSESE, the overall comparison will use Cox proportional hazards models, while comparisons for individual components will use the Fine-Gray model to account for the competing risk of death. Models will be stratified by disease severity at study entry (*invasive mechanical ventilation or ECMO* vs. neither) if event numbers permit.

- Treatment groups will be compared for incidence of a composite of cardiovascular and thromboembolic events, a subset of the organ failure outcome (items b1, e2, e3, and f2 in Appendix B). Time-to-event methods will be used that take into account the competing risk of death (as described in section 3, using Aalen-Johansen estimates for the cumulative incidence functions, and Gray's and Fine-Gray's methods to compare treatment groups and estimate the sub-hazard ratio).

- **Subgroup analyses:** The impact of study arm on the primary safety outcome (composite of grade 3 or 4 events, SAEs, PSESEs, or death through Day 5) and other important safety outcomes will be assessed for subgroups defined by baseline characteristics, including demographics, duration of symptoms at enrollment, baseline classification of "home", clinical history and presentation (including disease severity at study entry), use of concomitant medications, and, if available, baseline levels of antibodies, antigen and viral RNA; tests for homogeneity of the treatment effect across subgroups will be carried out. Outcomes and methods for subgroup analyses are described in detail in section 8.5.

- Treatment groups will be compared for mean changes in laboratory test values from baseline to Day 3, and for incidence of grade 3 and 4 laboratory abnormalities at Day 3 (new abnormality or increase in grade). Laboratory tests include the basic metabolic panel (BMP), complete blood count (CBC) with differential, international normalizing ratio (INR), D-dimer, AST, ALT, and bilirubin. Statistical methods are described in section 3. Biomarkers will be log-transformed as needed.

  - For all participants, these biomarkers will be determined locally on Days 0, 1, and 2. Treatment groups will be compared for mean changes from baseline through

Day 2, overall using longitudinal models, and pointwise at Days 1 and 2 using ANCOVA models; models will be adjusted for baseline biomarker levels.
    - For participants who are in the ICU on Day 5, these biomarkers are also collected on Day 5. For this cohort, trajectories of mean biomarker values will be described, and treatment groups will be compared for mean changes in biomarker values from baseline through Day 5.
    - For renal function lab tests, participants who are on dialysis at study entry will be excluded.

- Pregnancy outcomes will be summarized.

Listings of SAEs, PSESEs, incident grade 3 and 4 AEs, UPs, SUSARs, and deaths (with cause of death) by treatment group will be provided at each DSMB meeting, with new events highlighted. The listings will include important baseline characteristics, such as age, sex, and level of respiratory support at study entry.

Further safety assessments may be considered.

Concomitant medication use is collected at baseline, and daily through Day 7; additionally, any use between Days 8 and 14, and use on Day 28 is reported. Vasopressor use is reported daily for Days 0-14. The following categories of concomitant medications will be summarized by treatment group:

- Corticosteroid use (daily for Days 0-7, any use through Day 14)
- Vasopressor use (daily for Days 0-14)
- Anticoagulation and antiplatelet therapy, by type
- COVID-19 treatments other than corticosteroids, by type
- Other summaries will be provided by request of the DSMB or study leadership.

7.2 Additional Safety Analyses for the Aviptadil Cohort

Limited safety data are available for aviptadil. Therefore, infusion reactions to aviptadil will be carefully monitored. Side effects of aviptadil observed in previous studies include:

- Hypotension with and without vasopressor
- Diarrhea
- Facial flushing
- Bradycardia
- Tachycardia

Data summaries that address these potential side effects during or within 2 hours after the infusion are described in section 6.1, under “Safety”.

In addition, the aviptadil and placebo groups will be compared for *hypotension associated with organ dysfunction* within the first 5 days after study entry. *Hypotension associated with organ dysfunction* is defined as hypotension plus concomitant or subsequent organ dysfunction. Organ dysfunction in this setting is a composite outcome consisting of items 5a-5f (excluding item 5b4) of the secondary outcome of *clinical organ failure and serious infections* (section 4.1.2 of the TESICO master protocol). The composite outcome will be assessed on Day 5, and compared between treatment groups using CMH tests, stratified by

disease severity at baseline; the ORs will be estimated with 95% CIs using logistic regression, adjusted for disease severity at baseline.

8 Efficacy Analyses

Analysis cohorts: Separate efficacy analyses will be conducted for the ***Aviptadil*** cohort (comparison of aviptadil versus placebo), and the ***Remdesivir*** cohort (comparison of remdesivir versus placebo). Interactions between aviptadil and remdesivir will be assessed in the ***Factorial*** cohort (assessment whether the effects of aviptadil and remdesivir are independent of each other, section 9).

Efficacy analyses will be by intention-to-treat, unless otherwise stated.

8.1 Primary Efficacy Endpoint

The **primary endpoint** of the trial is a 6-category ordinal outcome that assesses the participants' recovery status at Day 90, referred to "**recovery**". The six ordered categories of recovery are shown in Table 1 below; they consist of 3 ranked categories that describe the number of days alive, at home, and not receiving new supplemental oxygen ***at Day 90*** (77 or more consecutive days, 49–76 days, or 1–48 days) as well as an additional 3 categories for patients who are not recovered at Day 90. The ranking is from 1=best to 6=worst (death).

Table 1. Categories of the primary endpoint of *recovery at Day 90*

Category	Status at 90 days
1 (Best)	At home and off oxygen. No. of consecutive days at Day 90 ≥ 77
2	49-76
3	1-48
4	Not hospitalized AND either at home on oxygen OR not at home
5	Hospitalized for medical care OR in hospice care
6 (Worst)	Dead

Definition of *Home* for the primary endpoint:
According to the protocol, section 4.1, and consistent with the TICO protocol (NCT04501978), *Home* is defined as the level of residence or facility where the participant was residing prior to hospital admission leading to enrollment in this protocol.

Residence or facility groupings to define home are:

1) **Independent/community dwelling** with or without help, including house, apartment, undomiciled/homeless, shelter, or hotel

2) **Residential care facility** (e.g., assisted living facility, group home, other non-medical institutional setting)

3) **Other healthcare facility** (e.g., skilled nursing facility, acute rehab facility)

4) **Long-term acute care hospital** (hospital aimed at providing intensive, longer term acute care services, often for more than 28 days).

Lower (less intensive) level of residence or facility will also be considered as home. By definition, “home” cannot be a “short-term acute care” facility. Participants previously residing in a “long-term acute care” hospital recover when they return to the same or lower level of care.

Participants residing in a facility solely for public health or quarantine purposes will be considered as residing in the lowest level of required residence had these public health measures not been instated.

Since some patients may have been receiving supplemental oxygen before their COVID-19 illness, we define **new supplemental oxygen** as any supplemental oxygen in participants who were not receiving supplemental oxygen before their COVID-19 illness or an increase in supplemental oxygen above pre-COVID-19 baseline among patients who were receiving supplemental oxygen before their COVID-19.

The **“last-off” method** for assessing recovery will be used, as has been customary in the use of similar ordinal endpoints in ARDS trials for decades. According to the “last-off” method, periods of recovery that are followed by hospital re-admission, change from home to a higher level of care, or receipt of new supplemental oxygen will *not be counted* toward the number of days of recovery. In other words, only days between the last time the patient entered a recovered state (returned home, free of new supplemental oxygen), and Day 90 are counted as days of recovery.

8.2 Primary Analysis

Primary analysis
The investigational agent will be compared to the corresponding placebo group for *recovery at Day 90* by intention-to-treat. The primary analysis will use a proportional odds regression model to estimate a summary odds ratio (OR) for being in a better category in the investigational agent group compared with placebo; an OR > 1.0 will reflect a more favorable outcome for patients randomized to the investigational agent vs. placebo. The model will include a treatment indicator, and will be stratified by disease severity by including an indicator for receipt of *mechanical ventilation or ECMO* at enrollment.

Comments:

1. The primary endpoint can be ascertained only after 90 days of follow-up are completed (except in the case of death), since it is possible that a participant gets re-admitted to the hospital at any time after the initial discharge.
2. Ascertainment of the primary endpoint requires knowledge of the hospitalization status, type of residence (“home”), and oxygen use over time through Day 90. After the first hospital discharge, these will be assessed every 2 weeks (starting at Day 14, usually through phone contact).

3. At interim analyses, tests that compare treatment groups for the ordinal endpoint of "recovery at 90 days" need to account for the censoring of follow-up for all participants who have not yet completed 90 days. A novel statistical method combining proportional odds models with methods for competing risks is currently under review at a peer-reviewed journal (Tsiatis and Davidian 2021, personal communication). When published, this method will be implemented for interim analyses of the primary endpoint. The method requires that a certain proportion of participants has completed the 90 day follow-up, and such is not suitable for very early reviews.
4. **Missing data at the final analysis:** If the proportion of missing data is low and data are missing at random, the method by Tsiatis and Davidian (2021) described under item 3 will be applied. Prior to unblinding, the proportion, pattern, and reasons for missing data will be reviewed by the unblinded statisticians, and the method for treating missing data will be defined in cooperation with the blinded statisticians.

- The number and percentage of participants in each of the six categories on Day 90 will be tabulated, and the adjusted summary OR of the active versus control group will be estimated with a 95% CI, using a proportional odds model as described above.

  Sensitivity Analyses:
  - In addition to the adjusted summary OR, the unadjusted summary OR with 95% CI will be shown (estimated using a proportional odds model without adjustment for disease severity). In the case that the adjusted OR differs substantially from the unadjusted OR, the reason for the deviation will be explored.
  - The primary comparison will be repeated after excluding participants who did not receive any of the investigational agent/placebo (modified intention-to-treat).

- To supplement the overall summary odds ratio for the 6-category *recovery* outcome, each dichotomized definition of improvement that can be formulated from the components of the ordinal outcomes will be considered separately; for example, treatment groups will be compared for the proportions of participants in category 1 on Day 90; for the proportions in categories 1 or 2 ("best two categories"), in categories 1-3, etc. Proportions will be tabulated, and odds ratios for active versus control groups will be estimated with 2-sided 95% CIs using logistic regression models. These analyses need to be interpreted with caution, because they are not adjusted for inflation of type I error due to multiple comparisons.
  - The validity of the proportional odds assumption for the primary endpoint will be assessed by testing for heterogeneity in the log ORs (for the treatment effect) across the dichotomized cumulative ordered categories in the corresponding logistic regression model and across the stratification covariate (partial proportional odds model, test for "unequal slopes"), as described in section 3.

- **Subgroup analyses** will be carried out for the primary outcome. The goal is to determine whether the treatment effect differs across subgroups, and to aid the DSMB in considerations on whether there are safety concerns in specific subgroups. Principles for subgroup analyses are described in section 8.5; here, subgroup analyses are based on the proportional odds models. In particular, heterogeneity of the treatment effect by disease severity at baseline will be assessed.

8.3 Key Secondary Outcomes

The TESICO protocol identifies four key secondary outcomes (protocol section 11.2):

- **Mortality** through Day 90 is a key secondary outcome; analyses are described in section 7.1. In addition, interim monitoring boundaries are based on time to death (rather than the primary ordinal endpoint of *recovery*). Corresponding analyses are described in section 10.2.

- To supplement the separate analyses of *recovery at day 90* and *time to death*, a composite endpoint that considers the number of days at home off oxygen and the time to death (instead of just survival status at day 90) as well as the other categories of the primary ordinal recovery outcome will be analyzed jointly using the "**win ratio**" method.[8] (This analysis will be performed when the trial is completed).

  The win ratio will be calculated using the matched pairs method described in Pocock (2012).[8] Pairs will be formed by matching participants by disease severity (requiring invasive mechanical ventilation or ECMO), and by ranking the participants in each treatment group according to a risk score, described in section 13.1, and pairing the participants in groups A (here referring to the investigational drug) and B (referring to control) with equal ranks. Details are given below.

  - If both treatment groups have the same number of observations, the win ratio is calculated as follows:
    Step 1: Calculate the risk score for all participants, and order participants by the risk score in each treatment group. If needed, break ties at random. Each participant forms a "matched pair" with the participant of equal rank order in the other treatment group.
    Step 2: For each pair, determine whether the participant in group A wins, loses, or neither:
    a. Compare pairs for *time to death,* for all pairs where one or both participants died. If the participant in group A died, wins and losses are computed as follows:
       - If the matched participant in group B has longer follow-up, then A loses and B wins.
       - If the matched participant in group B has shorter follow-up and is alive at the censoring date, then neither group wins.

       Repeat for pairs where the participant in group B died.
    b. Compare remaining pairs for category 5 of the primary outcome, *hospitalized for medical care OR in hospice care*
    c. Compare remaining pairs for category 4 of the primary outcome, *not hospitalized AND either at home on oxygen OR not at home*
    d. Compare remaining pairs for *days off oxygen at home*.
       - If time at home off oxygen is longer for A, then A wins and B loses; vice versa for B.
       - If A achieved recovery (discharged home and off oxygen) at the latest follow-up date, and B was censored without reaching recovery before A reached recovery, then neither group wins; vice versa for B.
       - Otherwise, neither group wins.

Step 3: Calculate the win ratio as the number of wins in group A divided by the total number of pairs with a win or a loss in group A. Calculate the 95% CI for the win ratio and p-value as described in Pocock (2012).[8]

- If one treatment group has more participants than the other, select $|n_A - n_B|$ participants at random from the larger group and delete. Calculate the win ratio, 95% CI and p-value for the resulting matched pairs. Repeat the random selection of observations to delete 501 times (or more); identify the matched pairs data set that corresponds to the median win ratio; the final values of the win ratio, 95% CI and p-value are those calculated from this data set.
- If both treatment groups have the same number of observations, but some ranked risk scores are tied within a treatment group, a similar process may be used to repeat the random breaking of ties, with the final win ratio chosen as the median over repeated random tie breaks.

With this approach, time to death is first used to determine the winning group (i.e., longer time to death), then categories 4 and 5 followed by days off oxygen at home are used to determine the winning group: in this manner, the win ratio combines these conflicting outcomes into a composite while recognizing the importance of mortality.

- **Time to recovery** through Day 90, defined as *alive, at home, and off new supplemental oxygen*. Here, "new supplemental oxygen" is defined as supplemental oxygen above the level used prior to the COVID-19 infection. The cumulative incidence functions for recovery taking into account death as a competing risk will be estimated using the Aalen-Johansen method and compared using Gray's test with ρ=0. The recovery rate ratio will be estimated using a Fine-Gray regression model. The comparisons between treatment groups will be stratified by disease severity at study entry.
  - Per protocol, the recovery status in the primary ordinal outcome is defined using the "last-off" method, i.e., hospital re-admission, change from home to a higher level of care, or receipt of new supplemental oxygen changes the status to "not recovered". In contrast, the cited statistical methods for time-to-first-event analyses require that "recovery" is defined as an absorbent state, i.e., once a participant achieved "recovery", they will not revert to a different state later. Therefore, we will present two analyses:
    - Time to *first* being discharged from the hospital, at home, and off new oxygen.
    - Time to *last* being discharged from the hospital, at home, and off new oxygen; this corresponds to the "last-off" method. Results will have to be interpreted with caution, as the last-off method violates assumptions for the standard time-to-(first-)event analyses.
  - Time to first being *discharged from the hospital, at home and off oxygen for 14 consecutive days* will be analyzed to aid in the interpretation of "time to recovery".

- Status on a 3-category ordinal outcome that includes (a) recovered (alive, at home, and off new oxygen), (b) alive but not recovered, and (c) dead, assessed at Day 90. Here, *new oxygen* is defined as supplemental oxygen above the level that was used prior to COVID-19.
  Treatment groups will be compared using proportional odds models, stratified by disease severity.

8.4 Other Secondary Outcomes

The protocol defines a number of secondary endpoints in addition to the four key endpoints described in section 8.3 above. These analyses will be carried out for the final report. Selected secondary endpoints may also be analyzed for interim monitoring reports, to help evaluate the safety and efficacy of the investigational agent.

Below, the secondary outcomes from section 4.1.2 of the protocol are cited, with a short description of the analysis methods. For each outcome, the treatment groups will be compared by intention-to-treat, stratified by disease severity at study entry, as described in section 3 under "stratification".

- Time to discharge from the initial hospitalization. Treatment groups will be compared using time-to-event methods that take into account the competing risk of death, similar to the analyses for time to recovery described in section 8.3.
    - Hospital readmissions will be summarized using methods for recurrent events (i.e. those who are readmitted will re-enter the risk set).[9]
- *Hospital-free days to Day 90* (days alive outside of a short-term acute care hospital up to day 90). For this analysis, the "last-off" method will be used, i.e., days from the latest hospital discharge to day 90 will be counted. A person who dies within 90 days will be assigned a value of -1, consistent with the approach taken in many trials of intensive care-based interventions. We will present the median days by group and test the hypothesis of no difference between arms with a Wilcoxon rank sum test.
    - For interim analyses, only participants who have reached Day 90 (administrative follow-up for those who died) will be included, to avoid bias. Alternatively, the current follow-up time may be used, censored at the time point when the outcome status was last known for participants who were alive.
- A composite of death, clinical organ failure, or serious infection through Days 28 and 90 (see Appendix B). Analyses were described in section 7, under "PSESE".
- **Time to sustained recovery** through Day 90, defined as being discharged from the index hospitalization, followed by being alive and *home* for 14 consecutive days. (This is the primary endpoint in the ACTIV-3/INSIGHT 014/TICO protocol.) The analyses methods will take into account the competing risk of death, using the Aalen-Johansen method to estimate cumulative incidence functions, and Gray's test and the Fine-Gray method for treatment comparisons.
- Outcomes assessed in other treatment trials of COVID-19 for hospitalized participants in order to facilitate meta analyses and facilitate generation of norms, including an ordinal scale measuring the degree of oxygen support through Day 14, time to discharge from the initial hospitalization, and binary outcomes defined by worsening based on the worst 3 categories of the primary ordinal recovery outcome at day 90. We will try to match the analyses in the other trials, to get results that can be compared. These analyses will not be performed for interim reports to the DSMB, unless requested.

- A composite of cardiovascular events (outcomes listed in items b1, e2 and e3 in Appendix B) and thromboembolic events (item f2) through Day 28 and Day 90. Time to event methods will be used that take into account the competing risk of death, e.g., Gray's test to compare treatment groups.

8.5 Subgroup Analyses

As stated in the protocol, subgroup analyses for the primary efficacy outcome (recovery at Day 90), and for key safety outcomes (composite of *grade 3 and 4 AEs, SAEs, PSESEs, or death* through Day 5 and Day 28, composite of SAEs, PSESEs, or death through Day 90, time to hospitalization or death through Month 6, and for time to death) will be performed to determine whether and how the treatment effect (active versus control) differs qualitatively across various **subgroups defined at baseline**, and whether there are safety concerns in specific subgroups.

Key subgroup analysis are by disease severity at study entry and by study design stratum; other important subgroups include subgroups by age, by sex, by categories of oxygen support at baseline, by duration of symptoms prior to enrollment, by pre-existing conditions, and by use of concomitant medications (in particular, corticosteroid use).

Subgroup analyses will be performed by the following baseline factors:

- Study design stratum (defined in section 1.2)
  - for aviptadil, design stratum 1 versus strata 2 and 4
  - for remdesivir, design stratum 1 versus stratum 4
- Disease severity (*invasive mechanical ventilation or ECMO* versus neither).
- Oxygen requirement at baseline (HFNC, NIV, invasive mechanical ventilation without ECMO, ECMO)
- Duration of symptoms prior to enrollment
- Age
- Sex at birth
- Race/ethnicity
- Geographic location
- Residence (home) at the time COVID-19 symptoms developed
- Body mass index (BMI)
- Presence of chronic medical conditions. Only conditions with $\geq$ 5% prevalence will be considered.
  - Compromised immune function, defined as current use of antirejection medication after transplant; cytotoxic chemotherapy; treatment with biological medicine for autoimmune disease or cancer, HIV, or immunosuppressive disorder other than HIV.
  - Metabolic/vascular co-morbidities, defined as history of diabetes mellitus requiring treatment, a cerebrovascular event (thrombotic or hemorrhagic), heart failure, or an MI or other acute coronary syndrome
  - Hypertension with and without history of other metabolic and vascular co-morbidity (4 groups, as described in section 5)
  - Cancer
  - COPD
  - Asthma

  - Renal impairment or renal replacement therapy
  - Hepatic impairment
- Use of selected concomitant medications
  - Use of corticosteroids (recommended as SOC in this study), overall and by oxygen requirement at baseline
  - Use of antiplatelet/anticoagulant therapy (prophylactic heparin, intermediate or therapeutic heparin or other anticoagulant therapy, none)
  - Use of vasopressors
  - Use of IL-6 inhibitors or JAK inhibitors
- SARS-CoV-2 vaccination status at baseline, overall and by whether or not the immune function was impaired.
- ***In the Aviptadil cohort:*** Use of remdesivir at study entry or randomization to remdesivir in design stratum 1 (the 2x2 factorial).

  Comment: This subgroup analysis will provide information on whether the effect of aviptadil is independent of remdesivir. The assessment of independence in the subgroup analysis is not protected by randomization and would complement the assessment of independence in the factorial cohort, which is protected by randomization.
- ***In the Remdesivir cohort:*** Randomization to aviptadil.

When SARS-CoV-2 antibody and antigen levels in plasma and RNA levels from mid-turbinate nasal swabs (viral load) are available, subgroups will also be considered by upper respiratory SARS-CoV-2 viral load, by antibody level, by neutralizing antibody level, and by antigen level at baseline.

Subgroup analyses for the primary endpoint of recovery will use proportional odds models, stratified by disease severity at study entry (if the sample size permits). Summary ORs with 95% CIs comparing the investigational agent versus control will be estimated for each subgroup. Global tests for heterogeneity of the treatment effect across subgroups will be carried out, by adding the interaction between the subgroup indicator and the treatment group indicator to the model. In case the subgroups were formed by categorizing a continuous variable, the interaction term will be formed between the subgroup indicator and the continuous variable.

Subgroup analyses for the primary safety endpoint at Day 5 will use logistic regression, stratified by disease severity at study entry (if the sample size permits). Subgroup analyses for safety endpoints that are analyzed using time-to-event methods (those analyzed through Day 28 or longer) will use stratified Cox proportional hazards models, since death is part of the composite endpoints and not a competing risk. HRs will be estimated for each subgroup, and global tests of heterogeneity of the treatment effect will be carried out, as described above.

Additionally, subgroup analyses will be conducted for subgroups formed by a disease progression risk score at baseline. The construction of this risk score will be finalized later, see section 12.1.

Subgroup analyses will not be adjusted for multiple comparisons; they are supportive to the primary endpoint analyses. Subgroup analyses will be interpreted with caution due to limited power and uncontrolled type I error.

9 Assessment of Independence in the Factorial Cohort

Participants in the factorial cohort were randomized 1:1:1:1 to each of the four treatment combinations formed by aviptadil / matched placebo *and* remdesivir / matched placebo in a 2x2 factorial design. In this cohort, we will assess whether the effects of aviptadil and remdesivir are independent of each other (additive in the corresponding models). If there is evidence for an interaction effect (i.e., the effect of the two treatments is not independent), then the nature of the interaction will be investigated.

The presence of interactions between aviptadil and remdesivir will be investigated for the following outcomes:

- The composite of incident grade 3 or 4 AEs, SAEs, PSESEs, or death through Day 5 (**primary safety outcome**) and through Day 28.
  - For the final analyses, the test for an interaction between aviptadil and remdesivir at Day 5 will be performed using a logistic regression model that contains indicator variables for the aviptadil/ placebo main effect, the remdesivir/ placebo main effect, their interaction, and the indicator for disease severity at study entry (the stratification variable used in the main analysis). Similar for Day 28.
  - For interim analyses, the interaction tests will be performed using the corresponding Cox proportional hazards models.

- **Time to death**: The test for an interaction between aviptadil and remdesivir will be performed using a Cox proportional hazards regression model that contains indicator variables for the aviptadil/ placebo main effect, the remdesivir/ placebo main effect, and their interaction; the model will be stratified by disease severity at study entry.

- **Primary efficacy outcome** (recovery at 90 Days): The test for an interaction between aviptadil and remdesivir will be performed using a proportional odds regression model that contains indicator variables for the aviptadil/ placebo main effect, the remdesivir/ placebo main effect, their interaction, and the indicator for disease severity at study entry (the stratification variable used in the main analysis). The interaction test for the primary efficacy outcome will be performed after the 90-day follow-up is completed.

If there is evidence for an interaction between aviptadil and remdesivir ($p \leq 0.05$ for the interaction effect), then the four treatment groups will be described:

- Number and percent of participants with events in each group
- Kaplan-Meier curves for the time-to-event outcomes.

Comment: For the aviptadil versus placebo comparison in the full aviptadil cohort, the presence of differential treatment effects across subgroups by remdesivir use (randomized to remdesivir or use of remdesivir at study entry versus neither) also provides information about possible interactions between aviptadil and remdesivir. However, participants are not randomized to these subgroups (other than those in the factorial cohort); therefore, comparisons across these subgroups may be confounded with other patient characteristics.

10 Interim Monitoring Guidelines for the DSMB

Each investigational agent (aviptadil and remdesivir) versus placebo comparison will be treated as a separate clinical trial; stopping boundaries will be derived to allow for multiple interim looks, but will not be additionally inflated to adjust for simultaneous analysis of two different agents.

The DSMB will be asked to recommend early termination or modification only when there is clear and substantial evidence of benefit or harm.

As a guideline, **early termination for benefit or futility** based on the primary endpoint (*recovery* at Day 90) is not recommended, as the endpoint requires follow-up through Day 90. Given the anticipated rapid enrollment, this endpoint would be infeasible to use for stopping boundaries for either efficacy or futility. In addition, given the relatively short follow-up period of 90 days for this target population, full follow-up for the primary and all secondary endpoints is considered important to evaluate the investigational agents to be studied. An exception to this guideline is if the DSMB believe there is clear and substantial evidence of a mortality benefit for an investigational agent.

10.1 Early Assessment of Safety

Because data on aviptadil are limited, the pace of enrollment will be initially restricted. A comprehensive safety review will be conducted after the first 40 participants have been enrolled for the aviptadil vs. placebo comparison and Day 5 data are available. An initial safety review may be conducted earlier, e.g., after approximately 20 participants are enrolled for the aviptadil/placebo comparison and have Day 5 safety data available.

After the initial safety review, weekly safety reviews will be conducted. At the discretion of the DSMB, the frequency and content of these (initially weekly) safety reports may be modified. The DSMB may also request additional data summaries.

Monitoring of safety will be based on the totality of evidence, as described in section 7.

10.2 Interim Monitoring Boundaries

The monitoring guidelines for this master protocol focus on asymmetric stopping boundaries for harm or efficacy **based on mortality**, and ongoing close monitoring of safety by the DSMB, based on the totality of evidence. The stopping boundaries are provided as a guideline to the DSMB.

- For assessment of **harm**, a Haybittle-Peto boundary using 2.5 standard deviations (SD) of the test statistic under the null hypothesis for the first 100 participants enrolled and 2.0 SD afterwards. Harm will be assessed using **all-cause mortality**, specifically using a hazard ratio from a proportional hazard model for the time to death associated with the investigational agent.

- For interim assessment of **efficacy**, a Haybittle-Peto boundary using a 3.0 SD threshold will be used after 100 participants have been enrolled and followed for at

least 5 days. Efficacy will be assessed using **all-cause mortality**, specifically using a hazard ratio from a proportional hazard model for the time to death associated with the investigational agent.

Comment: The proportional hazards models for time to death will include the treatment indicator; if event numbers permit, the model will be stratified by disease severity at study entry (receipt of mechanical ventilation or ECMO vs. neither) (as described in section 3 for time to event analyses).

At each full interim review after the first 100 participants have been enrolled and followed for at least 5 days, the following will be provided:

- Z-value of the test statistic comparing treatment groups for time to death, plotted over information time, and the asymmetric Haybittle-Peto boundaries for harm and superiority described above.

  In addition to the current value of the test statistic, the corresponding values of the test statistic at the previous reviews will be plotted over information time, (1) as presented at the previous DSMB meetings, and (2) re-calculated with current data (using the cut-dates of the previous reports).

  Comment: Assuming that Z>0 denotes superiority of the investigational treatment and Z<0 denotes harm, then the Haybittle-Peto boundary for harm would be crossed when Z<-2, and the boundary for efficacy would be crossed when Z>3.

- History of the estimated hazard ratios for time to death with 95% CIs and p-values at previous DSMB reviews, as presented, and recalculated with the current data (using the cut-date of the previous reports). The latter provides information on the influence of a possible time lag in the ascertainment of deaths.

10.3 Interim Monitoring for Futility

No interim monitoring for futility is planned.

11 Data Completeness and Study Conduct

According to the protocol, clinical data will be collected on eCRFs to be submitted on Days 0, 1, 2, 3, 5, 7, 14, 28, 42, 60, 75, 90 and 180; mortality and re-hospitalizations will be assessed through Day 180 (6 months), SAEs and PSESEs through Day 90. (SAEs that are related to the investigational agent and Unanticipated Problems [UPs] are reported through Day 180). After hospital discharge, visits may be conducted by phone. Plasma and serum for central testing and for storage will be collected at baseline, at Day 3 (for participants who are hospitalized) and on Day 5 (for participants who are at the ICU or equivalent). The data collection schedule is included in Appendices D and E of this SAP.

Data completeness and study conduct reports will be provided by treatment group (for the closed report) and pooled across treatment groups (for the open report). Data summaries for the infusion of the investigational agents on Days 0-2 are described in Section 6; several of

those reports are also relevant for monitoring study conduct and will be included in the open report or provided to study leadership, pooled across treatment groups.

The following data summaries will be provided to assess data completeness and study conduct:

- Number and percent of participants with protocol deviations, and type of protocol deviation
- Expected and observed number (% of expected) of participants who completed visits on Days 0-3, 5, 7, 14, 28, 42, 60, 75, and 90.
- ***For the Aviptadil cohort:*** Expected and observed number (% of expected) of participants with infusion forms on Days 0, 1, and 2.
- Length of follow-up: Median, IQR, range
- Number and percent of participants who withdrew consent or were (potentially) lost to follow-up (no contact and unknown vital status for 30+ days).
- If substantial numbers of participants are lost to follow-up (e.g., more than 10% of participants), Kaplan-Meier estimates for the cumulative proportion of participants who are lost to follow-up over time, by treatment group, will be provided (closed report only).
- Listing of participants who withdrew consent, including dates of randomization, disease severity stratum at baseline, receipt of study treatment, oxygen requirement and hospitalization status at last visit, date of withdrawal, and reason of withdrawal.
- **Ascertainment of the primary endpoint** (recovery at Day 90, a 6-category ordinal outcome) requires knowledge of the hospitalization status, type of residence (“home”), and oxygen use over time through Day 90. After the first hospital discharge, these will be assessed every 2 weeks (starting at Day 14, usually through phone contact). To assess the data completeness for ascertaining the primary endpoint, the expected and observed number (% of expected) of participants with *known status for the components* will be provided for Days 14, 28, 42, 60, 75, and 90:
  - vital status (also for Day 180);
  - status of hospitalization;
  - status of oxygen use;
  - if discharged, the status of the residence (“home” versus other).
- Collection of specimens: Expected and observed number (% of expected) of participants with specimens collected as specified by the protocol, by visit.

A visit counts as “expected” if the visit window has closed or the data have been received.

12 SARS-CoV-2 Antigen, Antibody, and RNA Levels

SARS-CoV-2 antigen and antibody levels will be determined centrally, from stored plasma samples, and thus may not be available at interim analyses. Similarly, SARS-CoV-2 RNA levels will be determined centrally from mid-turbinate nasal swabs. If data are available, analyses will be included in interim reports. Analysis plans will be developed when more information is available.

13 Exploratory Analyses

13.1 Disease Progression Risk Score

A disease progression risk score, calculated at baseline, will be used to form subgroups of participants with low or high predicted risk for subgroup analyses for safety and efficacy outcomes, and to pair participants for the win ratio analyses described in section 8.3.

The risk score will be developed for the final analyses, and the method will be specified at a later time. A possible method is to derive the risk score using a proportional odds model for the primary outcome of recovery (ordinal outcome), with baseline predictors including the oxygen requirement at study entry, age, sex, and indicator variables for the following risk factors: asthma/COPD, diabetes, CVD, heart failure, hypertension, immune impairment, and renal impairment. The risk score will be derived from the pooled data for the investigational agent/placebo groups. Thus, the risk score will be specific to each investigational agent.

14 Unblinding of Treatment Comparisons

For any investigational agent, trial results will be unblinded when all participants have completed 90 days of follow-up; results may be unblinded earlier upon the recommendation of the DSMB if the sponsor and study leadership concur. In this case, trial results for the investigational agent will be unblinded and reported with available data through 90 days of follow-up. After that, data collection will continue through 180 days as outlined in the data collection plan.

While the trial is ongoing, access to any data summaries by treatment group (investigational agent or control groups) will be restricted to the members of the DSMB, the DSMB's Executive Secretary, and the unblinded statisticians.

When the trial for an investigational agent is concluded, data for the investigational agent and the corresponding pooled control group will be unblinded and provided to the study team.

15 Distribution of Reports

- Open report: ACTIV-3b leadership team; DAIDS Medical Officer; selected NIAID staff; representatives of the companies; and all recipients of the unblinded closed report. After the DSMB meeting, the open report and the DSMB summary statement will be posted to the trial's web site, open to all investigators.
- Closed report: DSMB members, Executive Secretary of the DSMB, unblinded statisticians.
- Web reports (accessible by all investigators and study staff):
    - Enrollment summaries by site and over time (updated daily)
    - Baseline characteristics
    - Selected summary measures on data quality and study conduct (pooled across treatment groups).

- Additionally, selected summary measures on study conduct will be provided to study leadership upon request (pooled across treatment groups).

16 References

1. Zhao W. Mass weighted urn design--A new randomization algorithm for unequal allocations. Contemp Clin Trials 2015;43:209-16.
2. McCullagh P. Regression Models for Ordinal Data. Journal of the Royal Statistical Society Series B (Methodological) 1980;42:109-42.
3. Aalen OO, Johansen S. An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations. Scandinavian Journal of Statistics 1978;5:141-50.
4. Gray RJ. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. Annals of Statistics 1988;16:1141-54.
5. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc 1999;94:496-509.
6. Zhou B, Latouche A, Rocha V, Fine J. Competing risks regression for stratified data. Biometrics 2011;67:661-70.
7. Division of AIDS (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events, Corrected Version 2.1. 2017. (Accessed August 14, 2020, at https://rsc.niaid.nih.gov/sites/default/files/daidsgradingcorrectedv21.pdf.)
8. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J 2012;33:176-82.
9. Wei LJ, Lin DY, Weissfeld L. Regression-Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. J Am Stat Assoc 1989;84:1065-73.

Appendix A. Categories of the Primary Outcome (Recovery)

The primary endpoint is a 6-category ordinal outcome that assesses participant recovery status at Day 90. The primary ordinal endpoint is referred to as ***recovery***.

Table A-1 Categories of the primary endpoint

Category	Status at 90 days
1 (Best)	At home and off oxygen. No. of consecutive days at Day 90 ≥ 77
2	49-76
3	1-48
4	Not hospitalized AND either at home on oxygen OR not at home
5	Hospitalized for medical care OR in hospice care
6 (Worst)	Dead

Home is defined as the level of residence or facility where the participant was residing prior to onset of COVID-19 leading to the hospital admission that led to enrollment in this protocol. Residence or facility groupings to define home are: 1) **Independent/community dwelling** with or without help, including house, apartment, undomiciled/homeless, shelter, or hotel; 2) **Residential care facility** (e.g., assisted living facility, group home, other non-medical institutional setting); 3) **Other healthcare facility** (e.g., skilled nursing facility, acute rehab facility); and 4) **Long-term acute care hospital** (hospital aimed at providing intensive, longer term acute care services, often for more than 28 days). Lower (less intensive) level of residence or facility will also be considered as home. By definition, “home” cannot be a “short-term acute care” facility. Participants previously residing in a “long-term acute care” hospital recover when they return to the same or lower level of care. Participants residing in a facility solely for public health or quarantine purposes will be considered as residing in the lowest level of required residence had these public health measures not been instated. If such patients are receiving new supplemental oxygen, they will not be classified as recovered.

The “last-off” method for assessing recovery at Day 90 will be used, i.e., in case a higher level of care is required after an initial discharge home, only days between the last time the patient entered a recovered state (returned home, free of new supplemental oxygen), and Day 90 are counted as days of recovery.

Appendix B. Definition of Clinical Organ Failure and Serious Infection

According to the protocol, section 4.1.2., *clinical organ failure* is defined by development of any one or more of the following clinical events (see PIM for criteria for what constitutes each of these conditions):

a. Worsening respiratory dysfunction
    1. Increase in the level of respiratory support from high-flow nasal cannula or non-invasive mechanical ventilation at baseline to mechanical ventilation or ECMO, or from invasive mechanical ventilation at baseline to ECMO.

b. Cardiac and vascular dysfunction:
    1. Myocardial infarction (MI)
    2. Myocarditis or pericarditis
    3. Congestive heart failure (CHF): new onset NYHA class III or IV, or worsening to class III or IV
    4. Hypotension treated with vasopressor therapy
    5. Atrial or ventricular tachyarrhythmias

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. Haematological 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).

Serious infection is defined as:

g. Serious infection:
    1. Intercurrent, at least probable, documented serious disease caused by an infection *other than* SARS-CoV-2, requiring antimicrobial administration and care within an acute-care hospital.

Appendix C. Protocol-specified Exempt Serious Events (PSESE)

Protocol-specified exempt serious events (PSESE) are defined in the TESICO protocol section 10.2.3. These events are usually of similar severity as SAEs, but are **not** reported as SAEs, **unless** the investigator considered that there was a reasonable possibility that the study intervention (blinded investigational agent/ placebo or study-supplied SOC treatment) caused the event.

- Death
- Stroke
- Meningitis
- Encephalitis
- Myelitis
- Myocardial infarction
- Myocarditis
- Pericarditis
- New onset of worsening of CHF (NYHA class 3 or 4)
- Arterial or deep vein thromboembolic events
- Renal dysfunction treated with renal replacement therapy
- Hepatic decompensation
- Neurologic dysfunction, including acute delirium and transient ischemic events
- Disseminated intravascular coagulation
- Major bleeding events
- Serious infections
- Worsening respiratory failure
- Hypotension treated with vasopressor therapy
- Atrial or ventricular arrhythmias

Comment: PSESEs include all events in the composite outcome of *organ failure or serious infections* (described in Appendix B above), plus *death*.

Appendix D. Safety Data Collection

Table D-1. Overview of Safety Data Collection (protocol version 2.0, section 10).

	During and at least 2 hrs after infusion (all days on which infusion occurs)	Day 0–7	Day 14	Day 28	Day 90
Infusion-related reactions and symptoms of any grade[a]	X				
All grade 3 and 4 clinical AEs (new or increased in severity to Grade 3/4)	X	X	X[b]	X[b]	
Protocol-specified exempt serious events (PSESEs)[c]	Collected through Day 90				
SAEs that are not PSESEs	Collected through Day 90				
Unanticipated problems	Collected through End of Subject Participation (Day 180)				
Hospital admissions and deaths	Collected through End of Subject Participation (Day 180)				
Any SAE related[d] to study intervention	Collected through End of Subject Participation (Day 180)				

[a] This includes reporting of AEs of any grade present on day 0, before the first infusion. This allows assessment of whether a given AE is new after infusion.
[b] Participants will be asked about all new relevant adverse events of Grade 3 or 4 which have occurred since the last data collection, up to that time point. On these visits, AEs of Grade 1 or 2 that are present on the day of the visit will also be collected.
[c] These are explained and defined in section 10.2.3 of the protocol, and Appendix B of this SAP.
[d] Relatedness determined as per protocol rules in protocol section 10.1.5.

Table D-2. Overview of Safety Data Collection for Remdesivir (stratum 3) (protocol version 2.0, Appendix H2, Table 1).

	Day 0–7	Day 14	Day 28	Day 90
All grade 3 and 4 clinical AEs (new or increased in severity to Grade 3/4)	X	X[a]	X[a]	
Protocol-specified exempt serious events (PSESEs)[b]	Collected through Day 90			
SAEs that are not PSESEs	Collected through Day 90			
Unanticipated problems	Collected through End of Subject Participation (Day 180)			
Hospital admissions and deaths	Collected through End of Subject Participation (Day 180)			
Any SAEs related[c] to study intervention	Collected through End of Subject Participation (Day 180)			

[a] Participants will be asked about all new relevant adverse events which have occurred since the last data collection, up to that time point. On these visits, AEs of Grade 1 or 2 that are present on the day of the visit will also be collected.
[b] PSESEs are collected on designated forms and consist of events most likely occurring due to the underlying disease. PSESEs are study endpoints and will be reviewed by the DSMB regularly, but will be "exempt" from additional collection and reporting as adverse events for safety. See section 10.2.3 of the master protocol for further details
[c] Relatedness determined as per protocol rules in section 10.

Table D-3. Hypotension AE grading (protocol version 2.0, section 10.1.4, Table 5)

AE GRADING	GRADE 1 MILD	GRADE 2 MODERATE	GRADE 3 SEVERE	GRADE 4 LIFE-THREATENING
SERIOUSNESS GUIDANCE*	N	N	N (usually)	Y
Hypotension criteia that apply to all assessments	No intervention or complication meeting criteria for higher grade.	IVF ≥500 mL **OR** low-dose vasopressor (e.g. <0.1 NE [or equivalent])	Moderate-dose vasopressor (e.g. ≥0.1 NE [or equivalent]) **OR** ≥2 vasopressors **OR** multiple interventions	Life-threatening or clinically significant complications **OR** *persistent* clinically significant deterioration.
Additional hypotension criteria for aviptadil/placebo infusion days	No infusion change for hypotension	Decrease infusion rate *for hypotension* **OR** pause infusion with resumption *for hypotension*	Study drug discontinued for day *for hypotension* **OR** study drug not given for day *for hypotension* **OR** study drug discontinued permanently *for hypotension*	No additional criteria

* Guidance provides suggested seriousness alignment with AE grade but does not overrule investigator judgment. In particular, the presence of critical illness influences the threshold for considering a given hypotension AE 'life-threatening' or an 'important medical event.' Evaluation of other factors, including the intensity of intervention required and the event's impact on the patient, are required to determine event seriousness.

Appendix E. Schedule of Assessments

Table E-1. Schedule of Assessments (protocol version 2.0, Appendix B)

	Screen or Day 0	**Day 0**	**Study Day**													
Day	**−1/0**[1]	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**14**	**28**	**42**	**60**	**75**	**90**	**180**
Acceptable deviation from day	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**+1**	**+2**	**+3**	**+3**	**+5**	**+5**	**+ 10**	**± 14**
ELIGIBILITY & BASELINE DATA																
Informed consent	X															
Baseline medical and social history	X															
Baseline concomitant medications	X															
Symptom-directed physical exam by the clinical team (includes vital signs)	X															
Nasal swab for virus detection and review SARS-CoV-2 test results	X															
Baseline study labs (CBC with differential, ferritin, CRP, BMP, INR, D-DIMER, AST, ALT, bilirubin)[2]	X															
Research sample storage (includes DNA and RNA at baseline among patients who consent to genetics)	X															
Urine pregnancy test or other documentation of pregnancy status	X															
STUDY INTERVENTION																
Randomization		X														
Study Drug/Placebo Administration[3]		X	X	X												
Assess infusion completion and adverse reactions[3]		X	X	X												

	Screen or Day 0	Day 0	Study Day													
Day	**−1/0**[1]	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**14**	**28**	**42**	**60**	**75**	**90**	**180**
Acceptable deviation from day	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**+1**	**+2**	**+3**	**+3**	**+5**	**+5**	**+ 10**	**± 14**
STUDY PROCEDURES																
Post-randomization concomitant medications		X	X	X	X	X	X	X	X	X[4]	X					
On-study labs (BMP, CBC with differential, INR, D-DIMER, AST, ALT, bilirubin)[2,5]		X	X	X												
Clinical labs (BMP, CBC with differential, INR, D-DIMER, AST, ALT, bilirubin)[5,6]					X[7]		X[8]									
Research sample storage (includes RNA at day 3 among patients who consent to genetics)[4]					X[7]		X[8]									
Vital signs[5]	X	X	X	X			X			X						
Hospitalization status					X		X		X	X	X	X	X	X	X	X
Changes in residence/facility										X	X	X	X	X	X	
Interim medical history									X	X	X	X	X	X	X[9]	X[9]
Oxygen support (for WHO/NIH/TICO ordinal outcome)	X	X	X	X	X	X	X	X	X	X[4]						
Clinical AEs of grade 3 and 4 severity		X	X	X	X	X	X	X	X	X	X					
Clinical AEs of any grade on day indicated										X	X					
SAEs and PSESEs		Report through 90 days														
SAEs related to study interventions		Report as they occur														
Unanticipated problems		Report as they occur														
Deaths and readmissions		Report as they occur														
Hospitalization Summary		Report upon hospital discharge														

[1] Screening must be performed within 24 hours of randomization.
[2] These laboratory evaluations will only be performed as study procedures if they are unavailable clinically on that study day
[3] Duration of study drug administration may vary by investigational agent; the sample provided here is for 3 successive days. Where the duration of study drug administration

varies from this schedule, the duration will be specified in the relevant agent-specific Appendix.

[4] The Day 14 visit will record values for Days 8–14.

[5] These will be not be collected after hospital discharge.

[6] These laboratory assessments will only include clinically available results

[7] It is acceptable to perform the Day 3 draw on Day 4.

[8] The Day-5 draw will occur only among patients who remain in the intensive care unit (ICU) or equivalent. It is acceptable to perform the Day 5 draw on Day 5±1, but the Day 3 and Day 5 draws can not both be performed on Day 4.

[9] Includes telephone administration of the Euro-QOL-5D-5L instrument.

Appendix F. List of Acronyms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
ACTT	Adaptive COVID-19 Treatment Trial
ADE	Antibody-dependent enhancement
AE	Adverse event
ARDS	Acute respiratory distress syndrome
BMP	Basic metabolic panel
CHF	Congestive heart failure
CI	Confidence interval
CIF	Cumulative incidence curve
CMH	Cochran-Mantel-Haenszel [test]
COVID-19	Coronavirus-Induced Disease 2019
CVA	Cerebrovascular accident
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal membrane oxygenation
EU	European Union
FDA	Food and Drug Administration (US)
FiO2	Fraction of inspired oxygen
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
GEE	Generalized estimating equations
GMT	Geometric mean titer
HFNC	High-flow nasal cannula oxygen
HR	Hazard ratio
ICC	International Coordinating Center
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
ICU	Intensive care unit
IgG	Immunoglobulin G
IL-6	Interleukin 6
INSIGHT	International Network for Strategic Initiatives in Global HIV Trials
IQR	Interquartile range
IRB	Institutional Review Board
ITT	Intention-to-treat
IV	Intravenous
nMAb	Neutralizing monoclonal antibodies
MedDRA	Medical Dictionary for Regulatory Activities
MI	Myocardial infarction
mITT	modified intention-to-treat
mL	Milliliter
MV	(Invasive) mechanical ventilation
NE	Norepinephrine equivalent (dose)
NEW	National Early Warning [score]
NIAID	National Institute of Allergy and Infectious Diseases, NIH (US)
NIH	National Institutes of Health (US)
NIHSS	National Institutes of Health Stroke Scale/Score
NIV	Non-invasive ventilation
NYHA	New York Heart Association

nMAb	Neutralizing Monoclonal Antibodies
OR	Odds ratio
PCR	Polymerase chain reaction
PIM	Protocol Instruction Manual
PT	Preferred term
PSESE	Protocol-specified exempt serious event
RNA	Ribonucleic acid
RR	Rate ratio
RRR	Recovery rate ratio
SAE	Serious adverse event
SARS-CoV-1	Severe acute respiratory syndrome coronavirus 1
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SAP	Statistical analysis plan
SOC	Standard of care
SpO2	Oxygen saturation by pulse oxymeter
SUSAR	Suspected unexpected serious adverse reaction
UMN	University of Minnesota
UP	Unanticipated problem
U.S.	United States of America
VIP	Vasoactive Intestinal Peptide
WHO	World Health Organization