Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Version 1.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Protocol Version 1.0

ClinicalTrials.gov Identifier: pending

July 29, 2020

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**
VERSION HISTORY **4**
GLOSSARY OF TERMS **5**
1 INTRODUCTION **6**
1.1 Purpose **6**
1.2 Key Updates to the SAP **6**
2 STUDY OVERVIEW **6**
2.1 Study Design **6**
2.2 Randomization Process **6**
2.3 Study Objectives **7**
2.3.1 Primary Objectives 7
2.3.2 Secondary Objectives 8
2.3.3 Exploratory Objectives 8
2.4 Overview of Sample Size Considerations **10**
2.4.1 Phase II 10
2.4.2 Phase III 10
2.5 Overview of Formal Interim Monitoring **11**
2.5.1 Phase II 11
2.5.2 Phase III 12
2.6 Graduation to Phase III **13**
3 OUTCOME MEASURES **14**
3.1 Primary Outcome Measures Phase II **14**
3.2 Primary Outcome Measures Phase III **14**
3.3 Secondary Outcome Measures **15**
3.4 Other Outcome Measures **18**
4 STATISTICAL PRINCIPLES **19**
4.1 General Considerations **19**
5 ANALYSIS APPROACHES **20**
5.1 Analyses of the Primary Objectives **20**
5.1.1 Primary Safety (Phase II and III) 21
5.1.2 Primary Clinical Symptoms (Phase II) 21
5.1.3 Primary Virologic (Phase II) 23
5.1.4 Primary Efficacy (Phase III) 25
5.2 Analyses of Secondary Objectives **27**
5.2.1 Secondary Clinical Symptoms 28
5.2.2 Secondary Virology 34
5.3 Exploratory Analyses **36**

5.3.1 New SARS-CoV-2 among Household Contacts 36
5.3.2 Hospitalization Course 37
5.3.3 Exploratory Virology 37
5.4 Interim Analysis Considerations 37
5.4.1 Phase II to Phase III Graduation Criteria 37
5.4.2 Phase III Statistical Considerations 37

Version History

Version	Changes Made	Date Finalized
1	Original Version	July 29, 2020

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CSF	Critical Success Factor
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
TOC	Trial Oversight Committee

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (SAP) describes the proposed content and general framework for the interim and primary statistical analysis reports of the phase II and phase III investigations of ACTIV-2/A5401. The Primary SAP addresses the primary, secondary and a subset of exploratory objectives of the study and describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov and that will be included in primary manuscripts. The Primary SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in corresponding supplements to this SAP.

1.2 Key Updates to the SAP

N/A

2 Study Overview

2.1 Study Design

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. It includes a phase II evaluation, followed by a transition into a larger phase III evaluation of promising agents that 'graduate' from phase II. The trial has a randomized, blinded, controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient testing of new agents against placebo within the same trial infrastructure. The graduation criteria may be changed (adapted) as new agents are included in the study and so analyses supporting the recommendation to graduate or otherwise are described in a separate analysis plan.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through week 24 to capture long-term safety information, hospitalizations or death.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 7 days prior to study entry with no more than 10 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 48 hours of study entry.

2.2 Randomization Process

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is having a similar number of participants on a given investigational agent and on the

comparison group for that agent. The comparison group for a given investigational agent includes all participants who were concurrently randomized to a placebo, who were also eligible to have received that investigational agent.

To achieve this, eligible participants will be randomized in two steps. The first randomization will be to the Investigational Agent Group, and the second randomization will be to investigational agent or placebo within the Investigational Agent Group they were assigned in the first randomization. Participants may be randomized to investigational agents that are in phase II evaluation or to agents that are in phase III evaluation.

For a given participant, the first randomization will occur at an equal ratio (e.g., 1:1, 1:1:1, …) with the ratio determined by the number (n) of investigational agents the participant is eligible to receive (if eligible for only one agent, then the participant would be assigned to the one appropriate Investigational Agent Group). For example, if there were three investigational agents and a participant was only eligible to receive two of the three (so n=2), the ratio used for their first randomization would be 1:1.

The second randomization will occur at a ratio of n:1, which is dependent on the number of investigational agents a participant is eligible to receive. In the example where a participant was only eligible for two of three available investigational agents, the second randomization to investigational agent or placebo would occur at a 2:1 ratio.

Both randomization steps will be stratified (using blocked randomization) by (1) time from symptom onset (≤ 5 days vs > 5 days), and (2) risk of progression to severe COVID-19 (‘high’ vs ‘low’), ‘high’ defined as a person with age ≥ 55 years or having a least one of several protocol-specified comorbidities.

Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from the study protocol; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed in supplementary analysis plans. In addition, exploratory objectives 1, 4, and 12 will also be addressed in this SAP; however, other exploratory objectives will be addressed in subsequent analysis plans.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with undetectable nasopharyngeal (NP) SARS-CoV-2 RNA at study days 3, 7, 14, 21, and 28 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces SARS-CoV-2 detection or levels of RNA in nasal swabs [Protocol Objective 1.2.3].

4) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.4].

5) Phase II: To evaluate differences in SARS-CoV-2 RNA levels in NP swabs between the investigational agent versus placebo treatment groups and among subgroups of the population and risk groups defined by age and comorbidities [Protocol Objective 1.2.5].

6) Phase II: To determine if the investigational agent reduces SARS-CoV-2 detection or levels of RNA in saliva and nasal swabs [Protocol Objective 1.2.6].

7) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.7].

8) Phase III: To evaluate differences in symptom duration between the investigational agent versus placebo treatment groups among subgroups of the population, and risk groups defined by age and comorbidities [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study week 24 [Protocol Objective 1.2.9].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes across the study population, notably sex, time from symptom onset to start of investigational agent, race/ethnicity, and risk groups defined by age and comorbidities [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course once a participant requires hospitalization [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the association between zinc and vitamin D levels and clinical outcomes and response to agents [Protocol Objective 1.3.11].

12) Phase II: To explore the impact of investigational agents on SARS-CoV-2 viremia, i.e., detection or level of SARS-CoV-2 RNA in the blood [Protocol Objective 1.3.12].

13) Phase II: To explore if self-collected nasal swabs and saliva correlate with the frequency of detection and levels of SARS-CoV-2 RNA in site-collected NP swabs [Protocol Objective 1.3.13].

2.4 Overview of Sample Size Considerations

The following is adapted from the protocol; further details on the assumptions and sample size calculation are provided in protocol section 10.4.

2.4.1 Phase II

For each investigational agent in phase II, the proposed sample size in 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with undetectable virus in the investigational agent group vs concurrent placebo group, regardless of the assumed percent undetectable in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

With respect to symptom duration, assuming 100 participants in each group will provide study diary data, the study will have 91% power to show a 20% relative reduction in median duration of symptoms, assuming:

- Log-10 Durations are normally distributed with 0.2 standard deviation;
- Wilcoxon rank sum test with two-sided 5% Type I error rate.

2.4.2 Phase III

For each investigational agent in phase III, the proposed sample size is 2000 participants consisting of 1000 participants who receive that agent and 1000 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size has been chosen to provide 88.7% power to detect a relative reduction of 33.3% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-

binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring during each phase of the study; additional details on monitoring can be found in protocol section 10.5. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to investigational agent or placebo or if two participants experience a Grade 4 AE deemed related to investigational agent or placebo, enrollment to the investigational agent or placebo group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

There will be an interim analysis of a given investigational agent when 50% of participants (i.e., 110 of the 220 for a given investigational agent group) have completed the day 14 evaluation and all data (including virology) is available in the database. This review will include analyses of interim safety and will evaluate the activity of the investigational agent via assessment of graduation criteria; see section 5.4.1 for details on graduation rules.

At this early review, if activity data support graduation to phase III and there are no safety concerns, then the DSMB may recommend to continue enrollment of participants into phase III without a pause at the end of phase II enrollment (i.e., continue enrollment while results from complete phase II follow-up are still pending). However, at this early review, if activity data do not yet support graduation, then enrollment will be paused at the end of phase II enrollment (i.e., no enrollment into phase III), until a review of complete phase II results, through day 28, occurs.

Regardless if enrollment to phase III is paused, the DSMB will also review results from complete phase II follow-up once all participants (n=220) have completed the day 28 evaluation. For investigational agents that have not graduated to phase III, if these results indicate that graduation criteria have been met and there are no safety or resistance concerns, then the DSMB may recommend continuation of the study into the phase III period of evaluation.

At the interim reviews, recommendations for graduation will depend on an acceptable safety profile. This will largely be based on differences in the frequency of Grade 3 or 4 AEs between participants receiving the investigational agent and placebo.

2.5.2 Phase III

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent). By-stratum summaries will also be reviewed.

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III, corresponding to 25%, 50%, and 75% of the expected maximal efficacy information of the trial. An additional early interim efficacy analysis will also be conducted at the end of phase II, which will be considered in calculating Type I error spending, though the total error spent at this analysis will be negligible given the early timing (i.e., ~10% of the expected information for a comparison of a given investigational agent vs placebo).

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo/control group and 10% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 33.3%), with 1000 participants per group, this corresponds to 250 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The earlier of when approximately 500 participants from the two groups combined (including phase II, 25% of the 2000) have been followed for the primary outcome assessed at day 28, or when approximately 62 participants in the two groups combined have been hospitalized/died (i.e. 25% of the expected 250 participants hospitalized/died);
- The earlier of when approximately 1000 participants from the two groups combined (50% of the 2000) have been followed for the primary outcome assessed at day 28, or when approximately 125 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 1500 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 187 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death, or for differences in

the primary outcome within strata. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome, or for those in one of the risk strata (e.g. high-risk participants or those treated closer to symptom onset). In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo/control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

Additional details on interim monitoring are provided in protocol section 10.5.

2.6 Graduation to Phase III

During the phase II period of the study, the DSMB will review interim safety and efficacy data to provide recommendations to the TOC via NIAID as to whether an investigational agent should graduate to phase III. The TOC will review DSMB recommendations, and may consider other secondary outcomes (e.g. dynamics of virologic measures and symptoms over time, or any evidence of viral rebound) in the decision to graduate an investigational agent from phase II to phase III.

The TOC will also consult with the company that owns the investigational agent, to determine the graduation decision. An independent, unblinded, group from the company will receive and review day 28 analysis data from the phase II comparisons of the investigational agent. The independent group will assist the company in deciding if the investigational agent should graduate to phase III and/or chose the dose of the phase III investigational agent. Based on these discussions and in consultation with the company, the TOC will decide whether an investigational agent enters into phase III.

3 Outcome Measures

All outcome measures are copied from the protocol. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the last day on or before study day 28 when any symptoms scored as moderate or severe at study entry are still scored as moderate or severe (i.e., not mild or absent), or any symptoms scored as mild or absent at study entry are scored as mild or worse (i.e., not absent). The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue, headache, chills, nasal obstruction or congestion, nasal discharge (runny nose), nausea or vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1) moderate (2) and severe (3).

3) Virologic: Detection (detectable versus undetectable) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, 21, and 28.
   [For Primary Objective 3 and Secondary Objective 5]

3.2 Primary Outcome Measures Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
   [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
[Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2 and Secondary 8]

   Duration defined as the same as the primary phase II outcome.

5) Phase II and III: Duration of fever through day 28 defined as the last day in the participant's study diary on which a temperature greater than 37.8°C was recorded or a potentially antipyretic drug, such as acetaminophen or ibuprofen, was taken.
[Supportive of Primary Objective 2 and Secondary Objective 8]

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2 and Secondary Objective 8]

7) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or placebo, hospitalization, and death. [For Secondary Objective 1].

   Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

8) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or placebo. [For Secondary Objective 2]

9) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 7]

10) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [For Secondary Objective 7]

Virology

11) Phase II only: Level (quantitative) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, 21, and 28.
    [Supportive of Primary Objective 3 and Secondary Objective 5]

12) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from site-collected NP swabs at days 0, 3, 7, 14, 21 and 28. [Supportive of Primary Objective 3 and Secondary Objective 5]

13) Phase II and III: Detection (detectable versus undetectable) of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 1-14 in phase II, and at entry and days 3, 7, 10, 14, 21, and 28 in phase III.

14) Phase II and III: Level of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 1-14, 21 and 28 in phase II, and at entry and days 3, 7, 10, 14, 21, and 28 in phase III.

15) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from participant-collected nasal swabs daily at days 0-14 and at days 21 and 28. [For Secondary Objective 3]

16) Phase II only: Detection (detectable versus undetectable) of SARS-CoV-2 RNA from saliva at days 3, 7, 14, 21, and 28. [For Secondary Objective 6]

17) Phase II only: Level of SARS-CoV-2 RNA from saliva at days 3, 7, 14, 21, and 28. [For Secondary Objective 6]

18) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from saliva samples at days 0, 3, 7, 14, 21, and 28. [For Secondary Objective 6]

Efficacy

19) Phase II only: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo. [Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

20) Phase II and III: Death from any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo. [Supportive of Primary Objective 4]

21) Phase II and III: Death from any cause or hospitalization during the 24-week period from and including the day of the first dose of investigational agent or placebo. [For Secondary Objective 9, with follow-up beyond the day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

22) Phase II and III: Death from any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo. [For Secondary Objective 9, with follow-up beyond the day 28]

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:
   - Death
   - Hospitalized, on invasive mechanical ventilation or ECMO;
   - Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   - Hospitalized, requiring supplemental oxygen;
   - Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28.
   [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24. [For Exploratory Objective 4]

   Ordinal scale defined as:
   - Death
   - Hospitalized, on invasive mechanical ventilation or ECMO;
   - Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   - Hospitalized, requiring supplemental oxygen;
   - Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4]

13) Phase II and III: Detection (detectable versus undetectable) of SARS-CoV-2 RNA in blood through day 28. [For Secondary Objective 6]

    Blood collected at entry and days 7, 14, 21, and 28 in phase II, and at entry and day 28 in phase III.

14) Phase II and III: Level of SARS-CoV-2 RNA in blood through day 28.
[For Secondary Objective 6]

    Blood collected at entry and days 7, 14, 21, and 28 in phase II, and at entry and day 28 in phase III.

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group.

- Treated Population: All participants who could have been randomized to the given Investigational Agent Group and received any investigational agent/placebo (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus placebo. Exclusion of participants who are randomized, but who do not start their investigational agent/placebo should not introduce bias as the study is blinded. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to a placebo, who were also eligible to have received the investigational agent of interest. The comparison group will pool across all relevant placebo groups. For the primary analysis of each investigational agent, a supplemental

analysis will restrict the comparison group to include only participants who received the placebo for that investigational agent of interest.

Study visit windows for reporting are as defined in the protocol. Key study visits are Entry (Day 0), day 28, week 24, and the day of last dose of investigational agent/placebo (day X); day of last dose of investigational agent/placebo depends on the specific investigational agent dosing schedule and investigational agent or placebo, see relevant protocol appendix for details. Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Entry (Day 0):	First dose of investigational agent/placebo occurs.
Day X:	Last day of investigational agent/placebo. See protocol appendices for details on specific investigational agents.
Day 28:	Last day primary outcome may occur.
Week 24:	Last study visit.

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized and placebo-controlled; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

5 Analysis Approaches

5.1 Analyses of the Primary Objectives

Analysis Population

The analyses of the primary objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations, such as those who

start investigational agent or placebo outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis, but the protocol violations will be documented and described.

Note: According to the protocol, participants who are randomized but do not start investigational agent or placebo are not to be followed and will be replaced.

5.1.1 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

Sensitivity Analyses

Because some agents may be administered using injections or infusions and others will not be, the primary safety analyses will be repeated, but will exclude any occurrence of Grade 3 or higher local injection/infusion site reactions for investigational agents/placebos administered by injection or infusion.

Supportive Analyses

Secondary outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed separately as part of a supplementary analysis report (for week 24 outcomes) in the same manner as the primary safety outcomes.

5.1.2 Primary Clinical Symptoms (Phase II)

Analysis Approaches

Duration of symptoms will be summarized with descriptive statistics. Participant specific durations will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle

pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

The symptom duration is defined as the time (days) from start of investigational agent/placebo to the last day on or before day 28 when any symptoms are reported as at least moderate for those that were moderate or severe at study entry, or are reported as at least mild for those that were mild or absent at study entry.

To operationalize this, a duration will be calculated for each targeted symptom. The symptom duration outcome measure will be the maximum duration across the targeted symptoms. For symptoms that are absent at study entry that remain as absent through day 28, a duration of zero will be assigned; however, for symptoms that are absent at entry and emerge as mild, moderate, or severe, duration will be calculated as the number of days from study entry to the last day the symptom was scored as mild, moderate, or severe. For symptoms that are a mild at study entry, the duration will be calculated as the number of days from study entry to the last day the symptom was scored as mild, moderate, or severe. For symptoms that are moderate or severe, the duration will be calculated as the number of days from study entry to the last day the symptom was scored as moderate or severe. For symptoms that remit during the 28-day period, but then reoccur, the period of remission will be ignored in calculating the duration.

Special considerations are made for participants who are hospitalized or die on or before day 28. For participants who become hospitalized on or before day 28, all symptoms are assumed to be at least moderate during hospitalization (i.e., imputed in analysis), regardless if they were present at study entry or at the time of hospitalization. Programmatically, all symptoms will be imputed as '*moderate'* during hospitalization (starting from day of hospital admission through to day of hospital discharge or day 28, whichever is earliest). Participants who die on or before day 28 will be ranked as the worst outcome (i.e., longest duration) in these analyses. Programmatically, all participants who die will be assigned a duration of 29 days. Diary cards that are filled out during hospitalization will be ignored (as, per protocol, they are not required to be completed during hospitalization), and the algorithm outlined above (and in the protocol) will be used during the hospitalization period.

Diary cards that are missing for reasons other than hospitalization or death will be ignored in the primary analysis. This is equivalent to assuming that missing symptom scores were absent (for symptoms reported as absent or mild at entry) or mild/absent (for symptoms reported as moderate or severe at entry). Programmatically, missing symptoms for reasons other than hospitalization or death will be imputed as '*absent'*.

Sensitivity Analyses

(1) The duration of symptoms analyses will be repeated using different assumptions for symptom scores that are missing for reasons other than hospitalization or death. In this analysis, the missing symptoms will be imputed as having a symptom severity that would extend duration (i.e., those with moderate/severe at entry will have missing scores imputed as moderate/severe, and those with absent/mild will have missing scores imputed as at least mild). Programmatically, missing symptoms for reasons other than hospitalization or death will be imputed as '*moderate'*.

In the event that major differences in the interpretation of results are observed between the primary analysis and this sensitivity analysis, analysis methods such as multiple imputation or inverse probability of censoring weights (IPCW) may be considered.

(2) A strength of the symptom duration definition is that it recognizes the possibility that symptoms may resolve and then reappear or may improve and then worsen. A weakness, however, is that the duration could be classified as long because, for example, of the appearance of a single symptom after a period with no symptoms. To assess sensitivity of the interpretation of the results to this type of issue, the following analysis of duration will be done. In this analysis, duration of symptoms will be defined as the time (days) from start of investigational agent/placebo to the day before two successive days of improved symptoms. Improved symptoms is defined as having all symptoms that were scored moderate/severe at baseline be resolved to absent/mild and all symptoms that were scored absent/mild at baseline be resolved to absent. Participants who are alive on day 28 and did not have two such successive days of improved symptoms, but all symptoms met these criteria on day 28, will be assigned a duration of 27 days; otherwise they will be assigned a duration of 28 days. Participants who die on or before day 28 will be assigned a duration as for the primary outcome definition above.

5.1.3 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with undetectable SARS-CoV-2 RNA by NP swabs at each scheduled measurement time (entry and days 3, 7, 14, 21, and 28).

The proportion of participants with undetectable SARS-CoV-2 RNA will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. For each post-entry time point, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between groups, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. In this analysis, baseline SARS-CoV-2 RNA values below assay lower limit of quantification (LLoQ) will be imputed as half the distance from zero to the log-10 transformed LLoQ, i.e., the value used in the model will be ($\log_{10}$[LLoQ] / 2). It is not expected that a high proportion of results

will be < LLoQ, However, in the event that there is a non-negligible amount of censoring (defined as 10% or more of baseline results < LLoQ), an additional adjustment will be added to the model that will indicate whether the result was above or below assay quantification limit (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

A joint test of randomized arm across the time points (5 degrees of freedom) will also be assessed. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. The estimated adjusted relative risk of being undetectable (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with undetectable SARS-CoV-2 RNA at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level with handling of detected levels below the LLoQ as described above.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For sporadic missingness, participants with missing SARS-CoV-2 results will have their values imputed as undetectable if the preceding and succeeding results are undetectable (for day 28: if preceding result is undetectable), otherwise the results will be imputed as detectable.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as *detectable*.
    - For sporadic missingness, participants with missing SARS-CoV-2 results will have their values imputed as *undetectable* if the preceding and succeeding results are undetectable (for day 28: if preceding result is undetectable), otherwise the results will be imputed as *detectable*.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with undetectable levels will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<55 years and no comorbidities, ≥ 55 years or at least one comorbidity)
5) Age Group (<55, ≥55)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.4 Primary Efficacy (Phase III)

Analysis Approaches

The analysis of the primary efficacy outcome will compare the cumulative proportion of participants hospitalized or died (from any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. For analysis purposes, the integer scale will be used as the time scale, where study day 1 is considered day 1 and study day 28 is considered day 28; if an event occurs on day zero then event time will be set to 0.5 for analysis. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion

calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be antilogged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs concurrent placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons.

1) Evaluate the composite outcome of being hospitalized, dead, or lost to follow up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28.

2) Evaluate the impact of differential loss-to-follow-up.

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

3) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that differences are observed between the primary analysis and this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

Supportive Analyses

Secondary outcome 20 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (from any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary outcomes 21 and 22, which address secondary objective 1.2.9 from the protocol, evaluate the proportion of participants who are hospitalized or died through week 24, and the proportion who die (from any cause) through week 24. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have

their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days).

Secondary outcome 19 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are sufficient number of participants who died or were hospitalized. If not, descriptive summaries of the deaths and hospitalizations will be done in phase II.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the proportion of participants in a subgroup is low, or the number of events is low, descriptive summaries of the number of hospitalizations and deaths will be done. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<55 years and no comorbidities, ≥ 55 years or at least one comorbidity)
5) Age Group (<55, ≥55)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom duration outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be ignored in analysis (i.e., these symptoms are assumed to have not progressed at the time of missingness).

Duration of Fever

Duration of fever will be summarized with descriptive statistics, and will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The calculation of fever duration will take into consideration the temperature readings reported by the participants, as well as the reported use of any anti-pyretic medications.

The fever duration is defined as the time (days) from study entry to the last day on or before day 28 when a fever was reported (temperature greater than 37.8°C) or anti-pyretic medications were reported as being used.

Participants who never report a temperature greater than 37.8°C and never report use of anti-pyretic medications will be assigned a duration of fever of zero days. For the main analysis, special considerations will not be made for missing diary cards due to hospitalization or death (as it is possible that all fevers resolved prior to hospitalization or death). Missing fever evaluations on diary cards are assumed to be missing completely at random (MCAR) and will be ignored in these analyses. Programmatically, missing fever evaluations on diary cards for any reasons will have fever imputed as “*no*”.

Return to Usual Health

Duration of time without self-reported return to usual health will be summarized with descriptive statistics, and will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the time (days) from start of treatment to the last day on or before day 28 that self-reported return to usual health was "*no*".

Participants who never report "no" after starting study treatment will be assigned a time of zero days.

Special considerations are made for participants who are hospitalized or die on or before day 28. For participants who are hospitalized, the diary card answer is imputed as "no" for the period of hospitalization. Programmatically, self-reported return to usual health will be imputed as '*no'* starting from day of hospital admission through to day of hospital discharge or day 28, whichever is earliest. Diary cards that are filled out during hospitalization will be ignored (as, per protocol, they are not required to be completed during hospitalization), and the algorithm outlined above (and in the protocol) will be used during the hospitalization period. Participants who die on or before day 28 will be ranked as the worst outcome (i.e., longest time without return to usual health) in these analyses. Programmatically, all participants who die will be assigned a time of 29 days.

Diary cards that are missing for reasons other than hospitalization or death will be ignored in the analysis, that is, it is assumed participants have returned to usual health. Programmatically, missing data for reasons other than hospitalization or death will be imputed as '*yes'*.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day).

For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to (number of daily diary cards between day 0 and day 28), in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
3) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
4) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1) and (2). Next, (3) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (4).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, 21, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms.

A joint test of randomized arm across the time points (5 degrees of freedom) will also be assessed. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below). The estimated adjusted relative risk of being undetectable (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

A composite test, simultaneously analyzing all measurements from all time points after starting study treatment, will also be performed (DeLong ER, DeLong DM, Clarke-Pearson DL. Biometrics. 1988 Sep 1:837-45). Missing data also are assumed to be missing completely at random (MCAR) and will be ignored in this analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Duration of fever

1) Repeat duration of fever analyses, however, special considerations will be given for missing diary cards due to hospitalization or death. For participants who are hospitalized on or before day 28, fever will be assumed to be present during hospitalization. Programmatically, fever will be imputed as "*yes*" during hospitalization (starting from day of hospital admission through to day of hospital discharge or day 28, whichever is earliest). Participants who die on or before day 28 will be ranked as the worst outcome (i.e., longest duration) in these analyses. Programmatically, all participants who die on or before day 28 will be assigned a duration of 29 days.
2) Repeat duration of fever analyses, but define duration of fever as the time from day 0 to the last day on or before day 28 when a fever was reported (temperature >37.8°C was recorded). This analysis will not make special considerations for participants who indicated using anti-pyretic medications (i.e., will not include the use of a potentially anti-pyretic drug in the definition of fever). In this sensitivity analysis, those who never report fever will be assigned duration of fever of zero days.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For sporadic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% (for day 28, if preceding result is ≥96%), otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For sporadic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% (for day 28, if preceding result is ≥96%), otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<55 years and no comorbidities, ≥ 55 years or at least one comorbidity)
5) Age Group (<55, ≥55)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2.2 Secondary Virology

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

Analysis Methods

Detection (Detectable vs Undetectable) of SARS-CoV-2 RNA

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with undetectable SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., anterior nasal swabs and saliva).

The proportion of participants with undetectable SARS-CoV-2 RNA will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. For each post-entry time point, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between groups, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. In this analysis, baseline SARS-CoV-2 RNA values below assay lower limit of quantification (LLoQ) will be imputed as half the distance from zero to the log-10 transformed LLoQ, i.e., the value used in the model will be ($\log_{10}$[LLoQ] / 2). It is not expected that a high proportion of results will be < LLoQ; however, in the event that there is a non-negligible amount of censoring (defined

as 10% or more of baseline results < LLoQ), an additional adjustment will be added to the model that will indicate whether the result was above or below assay quantification limit (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

In addition, a joint test of randomized arm across the time points will also be assessed (degrees of freedom determined by the number of time points specimens are collected, 5 degrees of freedom for NP and saliva in phase II, 16 degrees of freedom for nasal swabs in phase II, and 6 degrees of freedom for nasal swabs in phase III). This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness.

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., NP swabs, anterior nasal swabs and saliva).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below limited of quantification will be imputed as the lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

A composite test, simultaneously analyzing all post-entry time points will also be performed (DeLong ER, DeLong DM, Clarke-Pearson DL. Biometrics. 1988 Sep 1:837-45).

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

Levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs, anterior nasal swabs, and saliva will be analyzed using participant-specific AUCs; this will be done separately for each specimen type. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: $\max[\log_{10}(\text{LLoQ}), \log_{10}(\text{RNA})-\log_{10}(\text{LLoQ})]$, obtained at the scheduled measurement times between and including day 0 and day 28.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 28 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 28 days). If the

day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with undetectable SARS-CoV-2 RNA at Day 0.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For sporadic missingness, participants with missing SARS-CoV-2 results will have their values imputed as undetectable if the preceding and succeeding results are undetectable (for day 28: if preceding result is undetectable), otherwise the results will be imputed as detectable.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as *detectable*.
    - For sporadic missingness, participants with missing SARS-CoV-2 results will have their values imputed as *undetectable* if the preceding and succeeding results are undetectable (for day 28: if preceding result is undetectable), otherwise the results will be imputed as *detectable*.
    - For monotonic missingness, inverse probability weighted GEE will be used.

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will include all randomized participants who started an investigational agent or the concurrent placebo, and will restricted to participants who report that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent placebo

through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered ignorable in analysis. The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent placebo through day 28.

Analysis of new SARS-CoV-2 positivity, and SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent placebo who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. The proportion of participants with ICU admission, among those hospitalized, will be summarized with frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 24.

5.3.3 Exploratory Virology

The analysis of SARS-CoV-2 RNA in blood will be done in the same manner as the secondary analysis of SARS-CoV-2 RNA from saliva and nasal swabs. See section 5.2.2 for details.

5.4 Interim Analysis Considerations

5.4.1 Phase II to Phase III Graduation Criteria

Each investigational agent considered in phase II will be evaluated for graduation to phase III. Graduation will be based on there being a desired level of evidence of an effect of an investigational agent versus placebo on one or more virologic and clinical outcome measures, as well as consideration of safety. The plan for these analyses will be provided in a separate document.

5.4.2 Phase III Statistical Considerations

The DSMB will review interim data from the study including descriptive summaries of study conduct and adverse events, and efficacy analyses that contrast randomized arms. The primary outcome of death or hospitalization will be compared between groups using the statistical

methods outlined in this SAP; the secondary outcome of death from any cause will be also be compared between randomized arms. Interim efficacy analyses are planned when approximately 500, 1000, and 1500 participants have been followed for the primary outcome assessed at day 28, or earlier if the total number of hospitalizations or deaths is higher than anticipated (See SAP Section 2.5 or Protocol Section 10.5).

At each interim review, the stopping boundary for the primary analysis will be determined based on the proportion of planned maximum information that is available at the given review. The statistical information (Fisher's Information) at a given review will be calculated using the inverse of the variance (square of standard error) obtained from Greenwood's formula as part of the primary analysis. The maximum information will be pre-determined using the following formula (Tsiatis AA. Statistics in medicine. 2006 Oct 15;25(19):3236-44):

$$MI = \left\{\frac{\left(Z_{\alpha/2}+Z_{\beta}\right)^2}{\delta_A^2}\right\} * (Inflation\ Factor) = \frac{(1.96+1.28)^2}{\left\{\ln\left(\frac{0.10}{0.15}\right)\right\}^2} * 1.03 = 65.8.$$

As a sensitivity analysis, we will assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event one day after the date they were lost to follow up. If the interpretation of the results from the primary analysis and this sensitivity analysis are substantially different, then considerations of the potential impact of delayed ascertainment of the primary endpoint will be considered, using an approach suggested by a DSMB statistician (personal correspondence A.A. Tsiatis).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Version 2.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 1.0, Clarification Memos #1 and #2, and Letters of Amendment #1, #2, and #3

ClinicalTrials.gov Identifier: NCT04518410

January 19, 2021

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**
VERSION HISTORY **4**
GLOSSARY OF TERMS **5**
1 INTRODUCTION **7**
1.1 Purpose **7**
1.2 Key Updates to the SAP **7**
1.2.1 Version 2.0 7
2 STUDY OVERVIEW **8**
2.1 Study Design **8**
2.2 Randomization Process **8**
2.3 Study Objectives **9**
2.3.1 Primary Objectives 9
2.3.2 Secondary Objectives 10
2.3.3 Exploratory Objectives 10
2.4 Overview of Sample Size Considerations **11**
2.4.1 Phase II 11
2.4.2 Phase III 12
2.5 Overview of Formal Interim Monitoring **12**
2.5.1 Phase II 13
2.5.2 Phase III 13
2.6 Graduation to Phase III **15**
3 OUTCOME MEASURES **15**
3.1 Primary Outcome Measures Phase II **15**
3.2 Primary Outcome Measures Phase III **16**
3.3 Secondary Outcome Measures **16**
3.4 Other Outcome Measures **19**
4 STATISTICAL PRINCIPLES **21**
4.1 General Considerations **21**
5 ANALYSIS APPROACHES **23**
5.1 Analyses of the Primary Objectives **23**
5.1.1 Primary Safety (Phase II and III) 24
5.1.2 Primary Clinical Symptoms (Phase II) 24
5.1.3 Primary Virologic (Phase II) 26
5.1.4 Primary Efficacy (Phase III) 28
5.2 Analyses of Secondary Objectives **31**
5.2.1 Secondary Clinical Symptoms 31
5.2.2 Secondary Virology 37

5.3 Exploratory Analyses 40
5.3.1 New SARS-CoV-2 among Household Contacts 40
5.3.2 Hospitalization Course 41
5.3.3 Exploratory Virology 41
5.3.4 Resistance Mutations 41
5.4 Interim Analysis Considerations 41
5.4.1 Phase II to Phase III Graduation Criteria 41
5.4.2 Phase III Statistical Considerations 41

6 APPENDIX 1: STATISTICAL CONSIDERATIONS FOR LY3819253 43
6.1 Phase II 43
6.2 Phase III 43
6.3 Phase II: Exploratory Analysis Pooling Over the 7000mg and 700mg Doses 44

Version History

Version	Changes Made	Date Finalized
1	Original Version	July 29, 2020
2	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CSF	Critical Success Factor
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2

SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (SAP) describes the proposed content and general framework for the interim and primary statistical analysis reports of the phase II and phase III investigations of ACTIV-2/A5401. The Primary SAP addresses the primary, secondary and a subset of exploratory objectives of the study and describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov and that will be included in primary manuscripts. The Primary SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in corresponding supplements to this SAP.

1.2 Key Updates to the SAP

1.2.1 Version 2.0

The following revisions have been made to the SAP in version 2.0:

1) Changes to protocol via CMs and LOAs
    a. Updated fever duration outcome per CM #1
    b. Added Appendix #1 to address statistical considerations for the LY3819253 agent in phase III per LOA #3
    c. No changes to SAP based on CM #2 or LOAs #1 or #2
2) Fixed typos/errors and added clarifications
    a. Added AUC outcome for RNA from blood (plasma)
    b. Added analyses by time point for dichotomous virology and oxygen saturation outcome measures
    c. Clarified that time points with zero events excluded in joint hypothesis test of treatment effects for dichotomous virology and oxygen saturation outcome measures over time
    d. Clarified that “sporadic” missingness means “non-monotonic” for analysis of dichotomous virology and oxygen saturation outcome measures
    e. Clarified that duration of fever is calculated from start of investigational agent/placebo (instead of from study entry) to be consistent across all analyses
    f. Clarified calculation of rescaling of AUC for symptom severity outcome
    g. Clarified primary dichotomous oxygen saturation analysis will adjust for baseline and added supportive analysis that does not adjust for baseline, consistent with virology analyses
3) Virology imputation
    a. Clarified how imputation of virology results will be done given that result can be <LoD and/or <LLoQ, and some may be >ULoQ
    b. Corrected typo in calculation of virology AUC

4) Missing symptom and virology data
    a. Revised how missing symptom data (for reasons other than hospitalization or death) will be handled and clarified how to handle missing Day 0 symptom data
    b. Clarified how to handle missing diary cards at Day 0 in the AUC symptom severity analysis
    c. Updated imputation of missing virology or oxygen data at Day 28 to be considered “monotonic” missingness in dichotomous analysis
5) Removed DeLong, DeLong, Clarke-Pearson composite test for oxygen saturation and virology
6) Added details related to analysis windows
7) Added analyses for resistance mutations including adding baseline mutations as subgroup in primary analyses to LY3819253 appendix.

2 Study Overview

2.1 Study Design

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. It includes a phase II evaluation, followed by a transition into a larger phase III evaluation of promising agents that ‘graduate’ from phase II. The trial has a randomized, blinded, controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient testing of new agents against placebo within the same trial infrastructure. The graduation criteria may be changed (adapted) as new agents are included in the study and so analyses supporting the recommendation to graduate or otherwise are described in a separate analysis plan.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through week 24 to capture long-term safety information, hospitalizations or death.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 7 days prior to study entry with no more than 10 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 48 hours of study entry.

2.2 Randomization Process

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is having a similar number of participants on a given investigational agent and on the comparison group for that agent. The comparison group for a given investigational agent includes all participants who were concurrently randomized to a placebo, who were also eligible to have received that investigational agent.

To achieve this, eligible participants will be randomized in two steps. The first randomization will be to the Investigational Agent Group, and the second randomization will be to investigational agent or placebo within the Investigational Agent Group they were assigned in the first

randomization. Participants may be randomized to investigational agents that are in phase II evaluation or to agents that are in phase III evaluation.

For a given participant, the first randomization will occur at an equal ratio (e.g., 1:1, 1:1:1, …) with the ratio determined by the number (n) of investigational agents the participant is eligible to receive (if eligible for only one agent, then the participant would be assigned to the one appropriate Investigational Agent Group). For example, if there were three investigational agents and a participant was only eligible to receive two of the three (so n=2), the ratio used for their first randomization would be 1:1.

The second randomization will occur at a ratio of n:1, which is dependent on the number of investigational agents a participant is eligible to receive. In the example where a participant was only eligible for two of three available investigational agents, the second randomization to investigational agent or placebo would occur at a 2:1 ratio.

Both randomization steps will be stratified (using blocked randomization) by (1) time from symptom onset (≤ 5 days vs > 5 days), and (2) risk of progression to severe COVID-19 ('high' vs 'low'), 'high' defined as a person with age ≥ 55 years or having a least one of several protocol-specified comorbidities.

Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from the study protocol; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed in supplementary analysis plans. In addition, exploratory objectives 1, 4, and 12 will also be addressed in this SAP; however, other exploratory objectives will be addressed in subsequent analysis plans.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with undetectable nasopharyngeal (NP) SARS-CoV-2 RNA at study days 3, 7, 14, 21, and 28 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces SARS-CoV-2 detection or levels of RNA in nasal swabs [Protocol Objective 1.2.3].

4) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.4].

5) Phase II: To evaluate differences in SARS-CoV-2 RNA levels in NP swabs between the investigational agent versus placebo treatment groups and among subgroups of the population and risk groups defined by age and comorbidities [Protocol Objective 1.2.5].

6) Phase II: To determine if the investigational agent reduces SARS-CoV-2 detection or levels of RNA in saliva and nasal swabs [Protocol Objective 1.2.6].

7) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.7].

8) Phase III: To evaluate differences in symptom duration between the investigational agent versus placebo treatment groups among subgroups of the population, and risk groups defined by age and comorbidities [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study week 24 [Protocol Objective 1.2.9].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes across the study population, notably sex, time from symptom onset to start of investigational agent, race/ethnicity, and risk groups defined by age and comorbidities [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course once a participant requires hospitalization [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the association between zinc and vitamin D levels and clinical outcomes and response to agents [Protocol Objective 1.3.11].

12) Phase II: To explore the impact of investigational agents on SARS-CoV-2 viremia, i.e., detection or level of SARS-CoV-2 RNA in the blood [Protocol Objective 1.3.12].

13) Phase II: To explore if self-collected nasal swabs and saliva correlate with the frequency of detection and levels of SARS-CoV-2 RNA in site-collected NP swabs [Protocol Objective 1.3.13].

2.4 Overview of Sample Size Considerations

The following is adapted from the protocol; further details on the assumptions and sample size calculation are provided in protocol section 10.4.

2.4.1 Phase II

For each investigational agent in phase II, the proposed sample size in 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with undetectable virus in the investigational agent group vs concurrent placebo group, regardless of the assumed percent undetectable in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

With respect to symptom duration, assuming 100 participants in each group will provide study diary data, the study will have 91% power to show a 20% relative reduction in median duration of symptoms, assuming:

- Log-10 Durations are normally distributed with 0.2 standard deviation;
- Wilcoxon rank sum test with two-sided 5% Type I error rate.

2.4.2 Phase III

For each investigational agent in phase III, the proposed sample size is 2000 participants consisting of 1000 participants who receive that agent and 1000 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size has been chosen to provide 88.7% power to detect a relative reduction of 33.3% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring during each phase of the study; additional details on monitoring can be found in protocol section 10.5. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to investigational agent or placebo or if two participants experience a Grade 4 AE deemed related to investigational agent or placebo, enrollment to the investigational agent or placebo group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

There will be an interim analysis of a given investigational agent when 50% of participants (i.e., 110 of the 220 for a given investigational agent group) have completed the day 14 evaluation and all data (including virology) is available in the database. This review will include analyses of interim safety and will evaluate the activity of the investigational agent via assessment of graduation criteria; see section 5.4.1 for details on graduation rules.

At this early review, if activity data support graduation to phase III and there are no safety concerns, then the DSMB may recommend to continue enrollment of participants into phase III without a pause at the end of phase II enrollment (i.e., continue enrollment while results from complete phase II follow-up are still pending). However, at this early review, if activity data do not yet support graduation, then enrollment will be paused at the end of phase II enrollment (i.e., no enrollment into phase III), until a review of complete phase II results, through day 28, occurs.

Regardless if enrollment to phase III is paused, the DSMB will also review results from complete phase II follow-up once all participants (n=220) have completed the day 28 evaluation. For investigational agents that have not graduated to phase III, if these results indicate that graduation criteria have been met and there are no safety or resistance concerns, then the DSMB may recommend continuation of the study into the phase III period of evaluation.

At the interim reviews, recommendations for graduation will depend on an acceptable safety profile. This will largely be based on differences in the frequency of Grade 3 or 4 AEs between participants receiving the investigational agent and placebo.

2.5.2 Phase III

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent). By-stratum summaries will also be reviewed.

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III, corresponding to 25%, 50%, and 75% of the expected maximal efficacy information of the trial. An additional early interim efficacy analysis will also be conducted at the end of phase II, which will be considered in calculating Type I error spending, though the total error spent at this analysis will be negligible given the early timing (i.e., ~10% of the expected information for a comparison of a given investigational agent vs placebo).

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo/control group and 10% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 33.3%), with 1000 participants per group, this corresponds to 250 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The earlier of when approximately 500 participants from the two groups combined (including phase II, 25% of the 2000) have been followed for the primary outcome assessed at day 28, or when approximately 62 participants in the two groups combined have been hospitalized/died (i.e. 25% of the expected 250 participants hospitalized/died);
- The earlier of when approximately 1000 participants from the two groups combined (50% of the 2000) have been followed for the primary outcome assessed at day 28, or when approximately 125 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 1500 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 187 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death, or for differences in the primary outcome within strata. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome, or for those in one of the risk strata (e.g. high-risk participants or those treated closer to symptom onset). In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo/control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

Additional details on interim monitoring are provided in protocol section 10.5.

2.6 Graduation to Phase III

During the phase II period of the study, the DSMB will review interim safety and efficacy data to provide recommendations to the TOC via NIAID as to whether an investigational agent should graduate to phase III. The TOC will review DSMB recommendations, and may consider other secondary outcomes (e.g. dynamics of virologic measures and symptoms over time, or any evidence of viral rebound) in the decision to graduate an investigational agent from phase II to phase III.

The TOC will also consult with the company that owns the investigational agent, to determine the graduation decision. An independent, unblinded, group from the company will receive and review day 28 analysis data from the phase II comparisons of the investigational agent. The independent group will assist the company in deciding if the investigational agent should graduate to phase III and/or chose the dose of the phase III investigational agent. Based on these discussions and in consultation with the company, the TOC will decide whether an investigational agent enters into phase III.

3 Outcome Measures

All outcome measures are copied from the protocol. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the last day on or before study day 28 when any symptoms scored as moderate or severe at study entry are still scored as moderate or severe (i.e., not mild or absent), or any symptoms scored as mild or absent at study entry are scored as mild or worse (i.e., not absent). The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue, headache, chills, nasal obstruction or congestion, nasal discharge (runny nose), nausea or vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1) moderate (2) and severe (3).

3) Virologic: Detection (detectable versus undetectable) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, 21, and 28.
   [For Primary Objective 3 and Secondary Objective 5]

3.2 Primary Outcome Measures Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
   [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
   [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2 and Secondary 8]

Duration defined as the same as the primary phase II outcome.

5) Phase II and III: Duration of fever through day 28 defined as the last day in the participant's study diary on which a temperature ≥ 38°C was recorded
[Supportive of Primary Objective 2 and Secondary Objective 8]

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2 and Secondary Objective 8]

7) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or placebo, hospitalization, and death. [For Secondary Objective 1].

Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank

order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

8) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or placebo. [For Secondary Objective 2]

9) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 7]

10) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [For Secondary Objective 7]

Virology

11) Phase II only: Level (quantitative) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, 21, and 28.
[Supportive of Primary Objective 3 and Secondary Objective 5]

12) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from site-collected NP swabs at days 0, 3, 7, 14, 21 and 28. [Supportive of Primary Objective 3 and Secondary Objective 5]

13) Phase II and III: Detection (detectable versus undetectable) of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 1-14 in phase II, and at entry and days 3, 7, 10, 14, 21, and 28 in phase III.

14) Phase II and III: Level of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 1-14, 21 and 28 in phase II, and at entry and days 3, 7, 10, 14, 21, and 28 in phase III.

15) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from participant-collected nasal swabs daily at days 0-14 and at days 21 and 28. [For Secondary Objective 3]

16) Phase II only: Detection (detectable versus undetectable) of SARS-CoV-2 RNA from saliva at days 3, 7, 14, 21, and 28. [For Secondary Objective 6]

17) Phase II only: Level of SARS-CoV-2 RNA from saliva at days 3, 7, 14, 21, and 28.
[For Secondary Objective 6]

18) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from saliva samples at days 0, 3, 7, 14, 21, and 28. [For Secondary Objective 6]

Efficacy

19) Phase II only: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

Hospitalization is defined as the same as the primary phase III outcome.

20) Phase II and III: Death from any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

21) Phase II and III: Death from any cause or hospitalization during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[For Secondary Objective 9, with follow-up beyond the day 28]

Hospitalization is defined as the same as the primary phase III outcome.

22) Phase II and III: Death from any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[For Secondary Objective 9, with follow-up beyond the day 28]

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28.
[For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24. [For Exploratory Objective 4]

Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4]

13) Phase II and III: Detection (detectable versus undetectable) of SARS-CoV-2 RNA in blood through day 28. [For Secondary Objective 6]

Blood collected at entry and days 7, 14, 21, and 28 in phase II, and at entry and day 28 in phase III.

14) Phase II and III: Level of SARS-CoV-2 RNA in blood through day 28.
[For Secondary Objective 6]

Blood collected at entry and days 7, 14, 21, and 28 in phase II, and at entry and day 28 in phase III.

15) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time in blood. [For Secondary Objective 6]

    Blood collected at entry and days 7, 14, 21, and 28 in phase II.

16) Phase II and III: Emergence of any new resistance mutations after study entry. [For Exploratory Objective 6]

    New resistance mutations are mutations that were not present at entry that were observed after study entry.

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group.

- Treated Population: All participants who could have been randomized to the given Investigational Agent Group and received any investigational agent/placebo (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus placebo. Exclusion of participants who are randomized, but who do not start their investigational agent/placebo should not introduce bias as the study is blinded. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to a placebo, who were also eligible to have received the investigational agent of interest. The comparison group will pool across all relevant placebo groups. For the primary analysis of each investigational agent, a supplemental analysis will restrict the comparison group to include only participants who received the placebo for that investigational agent of interest.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis

window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	**Protocol Range (Days)**	**Analysis Range (Days)**	**Analysis Window (Days)**
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	2, 4	+/- 1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 17	+/- 3
Day 21	21, 25	18, 25	-3, +4
Day 28	28, 32	26, 38	-2, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (ex. vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24, and the day of last dose of investigational agent/placebo (day X); day of last dose of investigational agent/placebo depends on the specific investigational agent dosing schedule and investigational agent or placebo, see relevant protocol appendix for details. Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Entry (Day 0): First dose of investigational agent/placebo occurs.

Day X: Last day of investigational agent/placebo.

See protocol appendices for details on specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Last study visit.

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized and placebo-controlled; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ and above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLOQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

5 Analysis Approaches

5.1 Analyses of the Primary Objectives

Analysis Population

The analyses of the primary objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations, such as those who start investigational agent or placebo outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis, but the protocol violations will be documented and described.

Note: According to the protocol, participants who are randomized but do not start investigational agent or placebo are not to be followed and will be replaced.

5.1.1 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

Sensitivity Analyses

Because some agents may be administered using injections or infusions and others will not be, the primary safety analyses will be repeated, but will exclude any occurrence of Grade 3 or higher local injection/infusion site reactions for investigational agents/placebos administered by injection or infusion.

Supportive Analyses

Secondary outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed separately as part of a supplementary analysis report (for week 24 outcomes) in the same manner as the primary safety outcomes.

In addition, for all analyses outlined above (primary, sensitivity, and supportive), the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

5.1.2 Primary Clinical Symptoms (Phase II)

Analysis Approaches

Duration of symptoms will be summarized with descriptive statistics. Participant specific durations will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored

daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

The symptom duration is defined as the time (days) from start of investigational agent/placebo to the last day on or before day 28 when any symptoms are reported as at least moderate for those that were moderate or severe at study entry, or are reported as at least mild for those that were mild or absent at study entry.

To operationalize this, a duration will be calculated for each targeted symptom. The symptom duration outcome measure will be the maximum duration across the targeted symptoms. For symptoms that are absent at study entry that remain as absent through day 28, a duration of zero will be assigned; however, for symptoms that are absent at entry and emerge as mild, moderate, or severe, duration will be calculated as the number of days from study entry to the last day the symptom was scored as mild, moderate, or severe. For symptoms that are a mild at study entry, the duration will be calculated as the number of days from study entry to the last day the symptom was scored as mild, moderate, or severe. For symptoms that are moderate or severe, the duration will be calculated as the number of days from study entry to the last day the symptom was scored as moderate or severe. For symptoms that remit during the 28-day period, but then reoccur, the period of remission will be ignored in calculating the duration.

Special considerations are made for participants who are hospitalized or die on or before day 28. For participants who become hospitalized on or before day 28, all symptoms are assumed to be at least moderate during hospitalization (i.e., imputed in analysis), regardless if they were present at study entry or at the time of hospitalization. Programmatically, all symptoms will be imputed as '*severe'* during hospitalization (starting from day of hospital admission through to the day before the day of hospital discharge or to day 28, whichever is earliest). Participants who die on or before day 28 will be ranked as the worst outcome (i.e., longest duration) in these analyses. Programmatically, all participants who die will be assigned a duration of 29 days. Diary cards that are filled out during hospitalization (starting from day of admission to day before day of discharge) will be ignored (as, per protocol, they are not required to be completed during hospitalization), and the algorithm outlined above (and in the protocol) will be used during the hospitalization period.

Missing values for reasons other than hospitalization or death will be imputed using the following algorithmic approach (after taking account of hospitalization and death as described above):

1) Impute missing value on Day 0 as "absent". If also missing on Day 1 or for a sequence of consecutive days from Day 1 but with at least one score during follow-up, impute the missing values through to the first available score as '*moderate*' [symptom duration will therefore be at least as long as the duration of a sequence of missing values starting at Day 0]

2) For intermittent missingness during follow-up, impute as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of length of sequence

of missing values [this gives potentially longer times until symptom improvement or resolution if either of the preceding and succeeding values don't meet the criterion for improvement or resolution, but potentially shorter time if both the preceding and succeeding values meet the criteria].

3) For monotonic missingness through to Day 28 (i.e. a sequence of missing values through to and including Day 28 due to loss to follow-up or participant choice not to complete the diaries), impute as '*moderate*', hence assuming that the relevant criterion for improvement or resolution has not been met [this has the effect of lengthening the symptom duration].

Sensitivity Analyses

(1) The duration of symptoms analyses will be repeated using different assumptions for symptom scores that are missing for reasons other than hospitalization or death. In this analysis, the missing symptoms will be imputed as '*absent*' so having the effect of potentially shortening the symptom duration versus the imputation used in the primary analysis.

(2) A strength of the symptom duration definition is that it recognizes the possibility that symptoms may resolve and then reappear or may improve and then worsen. A weakness, however, is that the duration could be classified as long because, for example, of the appearance of a single symptom after a period with no symptoms. To assess sensitivity of the interpretation of the results to this type of issue, the following analysis of duration will be done. In this analysis, duration of symptoms will be defined as the time (days) from start of investigational agent/placebo to the day before two successive days of improved symptoms. Improved symptoms is defined as having all symptoms that were scored moderate/severe at baseline be resolved to absent/mild and all symptoms that were scored absent/mild at baseline be resolved to absent. Participants who are alive on day 28 and did not have two such successive days of improved symptoms, but all symptoms met these criteria on day 28, will be assigned a duration of 27 days; otherwise they will be assigned a duration of 28 days. Participants who die on or before day 28 will be assigned a duration as for the primary outcome definition above.

5.1.3 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with undetectable SARS-CoV-2 RNA by NP swabs at each scheduled measurement time (entry and days 3, 7, 14, 21, and 28).

The proportion of participants with undetectable SARS-CoV-2 RNA will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated

measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of being undetectable (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

In this analysis, baseline SARS-CoV-2 RNA values will be imputed as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible amount of censoring (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below assay lower quantification limit (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with undetectable SARS-CoV-2 RNA at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level with handling of detected levels below the LLoQ as described above.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as undetectable if the preceding and succeeding results are undetectable, otherwise the results will be imputed as detectable.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association.

1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).

3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as *detectable*.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as *undetectable* if the preceding and succeeding results are undetectable, otherwise the results will be imputed as *detectable*.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with undetectable levels will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<55 years and no comorbidities, ≥ 55 years or at least one comorbidity)
5) Age Group (<55, ≥55)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

    Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.4 Primary Efficacy (Phase III)

Analysis Approaches

The analysis of the primary efficacy outcome will compare the cumulative proportion of participants hospitalized or died (from any cause), from day 0 through day 28, between

randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. For analysis purposes, the integer scale will be used as the time scale, where study day 1 is considered day 1 and study day 28 is considered day 28; if an event occurs on day zero then event time will be set to 0.5 for analysis. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs concurrent placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analyses is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or lost to follow up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that differences are observed between the primary analysis and this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up.

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and using variables after starting study treatment to

determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary outcome 20 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (from any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary outcomes 21 and 22, which address secondary objective 1.2.9 from the protocol, evaluate the proportion of participants who are hospitalized or died through week 24, and the proportion who die (from any cause) through week 24. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days).

Secondary outcome 19 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are sufficient number of participants who died or were hospitalized. If not, descriptive summaries of the deaths and hospitalizations will be done in phase II.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the proportion of participants in a subgroup is low, or the number of events is low, descriptive summaries of the number of hospitalizations and deaths will be done. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) ‘Risk of Severe Disease’ Stratification (<55 years and no comorbidities, ≥ 55 years or at least one comorbidity)
5) Age Group (<55, ≥55)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom duration outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Duration of Fever

Duration of fever will be summarized with descriptive statistics, and will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The calculation of fever duration will take into consideration the temperature readings reported by the participants.

The fever duration is defined as the time (days) from start of investigational agent/placebo to the last day on or before day 28 when a fever was reported (temperature ≥ 38°C).

Participants who never report a temperature ≥ 38°C will be assigned a duration of fever of zero days. For the main analysis, special considerations will not be made for missing diary cards due to hospitalization or death (as it is possible that all fevers resolved prior to hospitalization or death). As fevers are expected to be very infrequent at study entry, missing fever evaluations on diary cards are assumed to be missing completely at random (MCAR) and will be ignored in these analyses. Programmatically, missing fever evaluations on diary cards for any reasons will have fever imputed as "*no*".

Return to Usual Health

Duration of time without self-reported return to usual health will be summarized with descriptive statistics, and will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the time (days) from start of treatment to the last day on or before day 28 that self-reported return to usual health was "*no*".

Participants who never report "no" after starting study treatment will be assigned a time of zero days.

Special considerations are made for participants who are hospitalized or die on or before day 28. For participants who are hospitalized, the diary card answer is imputed as "no" for the period of hospitalization. Programmatically, self-reported return to usual health will be imputed as '*no'* starting from day of hospital admission through to day of hospital discharge or day 28, whichever is earliest. Diary cards that are filled out during hospitalization will be ignored (as, per protocol, they are not required to be completed during hospitalization), and the algorithm outlined above (and in the protocol) will be used during the hospitalization period. Participants who die on or before day 28 will be ranked as the worst outcome (i.e., longest time without return to usual health) in these analyses. Programmatically, all participants who die will be assigned a time of 29 days.

Return to Health answers that are missing for reasons other than hospitalization or death will be imputed in the analysis using the worst of the succeeding and preceding values. Return to Health answers that are missing at Day 0 and in a sequence of values starting at Day 0 for reasons other than hospitalization and death will imputed as "no". Monotonic missing values through to Day 28 will be imputed as "no".

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;

3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, 21, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard

errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Duration of fever

1) Repeat duration of fever analyses, however, special considerations will be given for missing diary cards due to hospitalization or death. For participants who are hospitalized on or before day 28, fever will be assumed to be present during hospitalization. Programmatically, fever will be imputed as "*yes*" during hospitalization (starting from day of hospital admission through to day of hospital discharge or day 28, whichever is earliest). Participants who die on or before day 28 will be ranked as the worst outcome (i.e., longest duration) in these analyses. Programmatically, all participants who die on or before day 28 will be assigned a duration of 29 days.
2) Repeat duration of fever analyses, but define duration of fever as the time from day 0 to the last day on or before day 28 when a fever was reported (temperature ≥ 38°C was recorded) or anti-pyretic medications were reported as being used in the participant's diary. This analysis makes special considerations for participants who indicated using anti-pyretic medications (i.e., will include the use of a potentially anti-pyretic drug in the definition of fever). In this sensitivity analysis, those who never report fever and never report use of anti-pyretic medications will be assigned duration of fever of zero days.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who

are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<55 years and no comorbidities, ≥ 55 years or at least one comorbidity)
5) Age Group (<55, ≥55)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2.2 Secondary Virology

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT)

approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

Analysis Methods

Detection (Detectable vs Undetectable) of SARS-CoV-2 RNA

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with undetectable SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., anterior nasal swabs and saliva).

The proportion of participants with undetectable SARS-CoV-2 RNA will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of being undetectable (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In addition, a joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

In this analysis, baseline SARS-CoV-2 RNA values will be imputed as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible amount of censoring (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below assay lower quantification limit (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., NP swabs, anterior nasal swabs and saliva).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

Levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs, anterior nasal swabs, and saliva will be analyzed using participant-specific AUCs; this will be done separately for each specimen type. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: $\max[0, \log_{10}(\text{RNA})-\log_{10}(\text{LLoQ})]$, obtained at the scheduled measurement times between and including day 0 and day 28.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 28 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 28 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

<u>Sensitivity Analyses</u>

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with undetectable SARS-CoV-2 RNA at Day 0.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as undetectable if the preceding and succeeding results are undetectable, otherwise the results will be imputed as detectable.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as *detectable*.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as *undetectable* if the preceding and succeeding results are undetectable, otherwise the results will be imputed as *detectable*.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with undetectable levels will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will include all randomized participants who started an investigational agent or the concurrent placebo, and will restricted to participants who report that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent placebo through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered ignorable in analysis. The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent placebo through day 28.

Analysis of new SARS-CoV-2 positivity, and SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent placebo who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 24.

5.3.3 Exploratory Virology

The analysis of SARS-CoV-2 RNA in blood (plasma) will be done in the same manner as the secondary analysis of SARS-CoV-2 RNA from saliva and nasal swabs. See section 5.2.2 for details.

5.3.4 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices.

5.4 Interim Analysis Considerations

5.4.1 Phase II to Phase III Graduation Criteria

Each investigational agent considered in phase II will be evaluated for graduation to phase III. Graduation will be based on there being a desired level of evidence of an effect of an investigational agent versus placebo on one or more virologic and clinical outcome measures, as well as consideration of safety. The plan for these analyses will be provided in a separate document.

5.4.2 Phase III Statistical Considerations

The DSMB will review interim data from the study including descriptive summaries of study conduct and adverse events, and efficacy analyses that contrast randomized arms. The primary outcome of death or hospitalization will be compared between groups using the statistical methods outlined in this SAP; the secondary outcome of death from any cause will be also be compared between randomized arms. Interim efficacy analyses are planned when approximately

500, 1000, and 1500 participants have been followed for the primary outcome assessed at day 28, or earlier if the total number of hospitalizations or deaths is higher than anticipated (See SAP Section 2.5 or Protocol Section 10.5).

At each interim review, the stopping boundary for the primary analysis will be determined based on the proportion of planned maximum information that is available at the given review. The statistical information (Fisher's Information) at a given review will be calculated using the inverse of the variance (square of standard error) obtained from Greenwood's formula as part of the primary analysis. The maximum information will be pre-determined using the following formula (Tsiatis AA. Statistics in medicine. 2006 Oct 15;25(19):3236-44):

$$MI = \left\{\frac{\left(Z_{\alpha/2}+Z_{\beta}\right)^2}{\delta_A^2}\right\} * (Inflation\ Factor) = \frac{(1.96+1.28)^2}{\left\{\ln\left(\frac{0.10}{0.15}\right)\right\}^2} * 1.03 = 65.8.$$

As a sensitivity analysis, we will assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event one day after the date they were lost to follow up. If the interpretation of the results from the primary analysis and this sensitivity analysis are substantially different, then considerations of the potential impact of delayed ascertainment of the primary endpoint will be considered, using an approach suggested by a DSMB statistician (personal correspondence A.A. Tsiatis).

6 Appendix 1: Statistical Considerations for LY3819253

6.1 Phase II

Emergence of Resistance Mutations

Four potential escape mutations were pre-identified for LY3819253: E484K, E484Q, E490S, and S494P. The laboratory conducting the resistance testing has stated that a mutation will be considered as present for a particular NP sample if at least 20% of the viral population in that sample have the mutation.

The emergence of resistant mutations after starting treatment will be summarized in the following manner. Descriptive statistics (number and percent) will be use to describe the proportion of participants with quantifiable (≥ LLoQ) SARS-CoV-2 RNA at Day 0 from NP swabs. Among those with and without quantifiable RNA at Day 0 from NP swabs, descriptive statistics will summarize the number and percent with quantifiable RNA during follow up from NP, and whether the participants developed new resistance mutations during follow up. The number and percent of mutations present at Day 0 will also be summarized.

The proportion of participants who have a new resistance mutation after study entry will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Note: this analysis assumes that those who were <LLoQ at Day 0 and stayed <LLoQ, and those who were <LLoQ throughout follow-up, did not develop resistance.

A supportive analysis will be exclude participants who had all mutations present at Day 0, as they could not develop new resistance.

Phase II primary virology analyses will include an additional subgroup analysis: Presence of Resistance Mutations (yes, no).

6.2 Phase III

After fully enrolling phase II (approximately 110 participants receiving LY3819253 and 110 receiving placebo), the LY3819253 agent will move directly (without graduation analysis of phase II data) into phase III as an open-label, single-arm, evaluation. There is no randomization in phase III. Enrollment into phase III will continue until another investigational agent enters the study; at this point phase III evaluation of LY3819253 will close.

Because of the single arm nature of the phase III evaluation, all phase III analyses will be descriptive using the same definitions for outcome measures and handling of missing data as described in the SAP. No formal analysis comparing participants who enrolled in the single arm Phase III component of the study with participants who participated in Phase II will be undertaken. If summaries are done by subgroups, and if available, summaries will also be done by whether there is presence of resistance mutations (yes, no).

6.3 Phase II: Exploratory Analysis Pooling Over the 7000mg and 700mg Doses

Analyses of LY3819253 will be done separately by dose (7000mg and 700mg), since each is considered as a separate agent in ACTIV-2 per protocol. In support of these analyses, exploratory analyses that pool across doses may also be considered (pooled 7000mg+700mg active vs pooled placebos for 7000mg+700mg).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Version 3.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 4.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0 & 4.0)

ClinicalTrials.gov Identifier: NCT04518410

April 2, 2021

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **4**

GLOSSARY OF TERMS **5**

1 INTRODUCTION **6**
1.1 Purpose **6**
1.2 Version History of this SAP **6**

2 STUDY OVERVIEW **6**
2.1 Study Design **6**
2.2 Randomization Process **7**
2.3 Study Objectives **8**
2.3.1 Primary Objectives 8
2.3.2 Secondary Objectives 9
2.3.3 Exploratory Objectives 9
2.4 Overview of Sample Size Considerations **10**
2.4.1 Phase II 10
2.4.2 Phase III – Infused Agents 11
2.5 Overview of Formal Interim Monitoring **11**
2.5.1 Phase II 12
2.5.2 Phase III - Infused Agents 12
2.6 Graduation to Phase III – Infused Agents **14**

3 OUTCOME MEASURES **15**
3.1 Primary Outcome Measures: Phase III **15**
3.2 Primary Outcome Measures: Phase II **15**
3.3 Secondary Outcome Measures **16**
3.4 Other Outcome Measures **18**

4 STATISTICAL PRINCIPLES **20**
4.1 General Considerations **20**

5 ANALYSIS APPROACHES **22**
5.1 Analyses of the Primary Objectives **23**
5.1.2 Phase III Primary Objective for Efficacy 23
5.1.2 Primary Safety (Phase II and III) 26
5.1.3 Primary Clinical Symptoms (Phase II) 27
5.1.4 Primary Virologic (Phase II) 30
5.2 Analyses of Secondary Objectives **33**
5.2.1 Secondary Clinical Symptoms 33
5.2.2 Secondary Virology 38
5.3 Exploratory Analyses **41**

5.3.1 New SARS-CoV-2 among Household Contacts 41
5.3.2 Hospitalization Course 42
5.3.3 Exploratory Virology 42
5.3.4 Resistance Mutations 42
5.4 Interim Analysis Considerations 43
5.4.1 Phase II to Phase III Graduation Criteria 43
5.4.2 Phase III Statistical Considerations – Infused Agents 43

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 44

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 47
7.1 Secondary Outcome Measures 47
7.2 Analysis Approaches 47

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 48
8.1 Secondary Outcome Measures 48
8.2 Analysis Approaches 48

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 49
9.1 Secondary Outcome Measures 49
9.2 Analysis Approaches 49

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 50
10.1 Objectives 50
10.1.1 Secondary Objectives 50
10.1.2 Exploratory Objectives 50
10.2 Outcome Measures 50
10.2.1 Secondary Outcome Measures 50
10.2.2 Other Outcome Measures 50
10.3 Analysis Approaches 50

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 51
11.1 Objectives 51
11.1.1 Secondary Objectives 51
11.1.2 Exploratory Objectives 51
11.2 Outcome Measures 51
11.2.1 Secondary Outcome Measures 51
11.3 Analysis Approaches 51

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 52

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III investigations of ACTIV-2/A5401. This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP are being limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253. Version 3.0 of the SAP is developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and is not being used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applies to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents will be described in agent-specific supplements to the SAP.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 4.0.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents. For infused agents, enrollment is restricted to participants at 'higher' risk for progression to severe COVID-19. For non-infused agents, enrollment is open to participants at both 'higher' and 'lower' risk of progression to severe COVID-19. See protocol for definition of 'higher' risk.

For infused agents, the study begins with a phase II evaluation, followed by a transition into a larger phase III evaluation of promising agents that 'graduate' from phase II. For non-infused

agents, the same phase II study will be undertaken as for infused agents, however, the design of the phase III evaluation for non-infused agents will be developed in a later version of the protocol.

The trial has a randomized, blinded, controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient testing of new agents against placebo within the same trial infrastructure. The graduation criteria may be changed as new agents are included in the study and so analyses supporting the recommendation to graduate or otherwise are described in a separate analysis plan.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through at least week 24 to capture long-term safety information, hospitalizations or death. Study visits may be required beyond week 24, depending on the investigational agent.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 8 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

The randomization process was the same under protocol versions 2.0, 3.0 and 4.0, and is summarized in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of participants on a given investigational agent and on the comparison group for that agent. The comparison group for a given investigational agent includes all participants who were concurrently randomized to a placebo arm in the same study phase as the investigational agent of interest, and who were also eligible to have received that investigational agent.

To achieve having a similar number of participants on the active arm and in the pooled placebo comparison group for a given investigational agent, the randomization will occur in two steps.

The first randomization will be to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

The second randomization is to the (active) investigational agent or placebo for that agent within an Agent Group. For a given participant, the probability of assignment to the active agent or placebo in the second randomization depends on (1) the number of agents currently under investigation that the participant was eligible to receive, and (2) the current study phase of the Agent Group that the participant was assigned to in the first randomization. For a participant who

was assigned to an Agent Group under evaluation in phase 2, the randomization will occur at a ratio of n_2:1, where n_2 is the number of investigational agents the participant was eligible to receive that are currently under investigation in phase 2. Similarly, for a participant who was assigned to an Agent Group under evaluation in phase 3, the randomization will occur at a ratio of n_3:1, where n_3 is the number of investigational agents the participant was eligible to receive that are currently under investigation in phase 3. Here, n (the total number of investigational agents the participant was eligible to receive in the first randomization) is equal to the sum of n_2 and n_3 (i.e., $n=n_2+n_3$).

Both the first and second randomizations involve blocked stratified randomization. For non-infused agents, both the first and second randomizations are stratified by (1) time from symptom onset (≤ 5 days vs > 5 days), and (2) risk of progression to severe COVID-19 ('higher' vs 'lower'). For infused agents, both the first and second randomizations are only stratified by time from symptom onset (≤ 5 days vs > 5 days), as only 'higher' risk participants are eligible for infused agents. A participant is considered at 'higher' risk of progression to severe COVID-19 if they have a least one of several protocol-specified factors (see protocol for details).Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 4.0; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed in supplementary analysis plans. In addition, exploratory objectives 1, 4, and 12 will also be addressed in this SAP; however, other exploratory objectives will be addressed in subsequent analysis plans.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, 14, and 28 [Protocol Objective 1.1.3].

4) Phase III for infused agents only: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in nasal swabs [Protocol Objective 1.2.3].

4) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.4].

5) Phase II: To evaluate differences in SARS-CoV-2 RNA levels in NP swabs between the investigational agent versus placebo and among subgroups of the population and risk groups defined by age and comorbidities [Protocol Objective 1.2.5].
6) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To evaluate differences in symptom duration between the investigational agent versus placebo treatment groups among subgroups of the population, and risk groups defined by age and comorbidities [Protocol Objective 1.2.7].

8) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study week 24 [Protocol Objective 1.2.8].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes across the study population, notably sex, time from symptom onset to start of investigational agent, race/ethnicity, and risk groups defined by age and comorbidities [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course once a participant requires hospitalization [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phase II: To explore the impact of investigational agents on SARS-CoV-2 viremia, i.e., detection or level of SARS-CoV-2 RNA in the blood [Protocol Objective 1.3.11].

12) Phase II: To explore if levels of SARS-CoV-2 RNA in self-collected nasal swabs correlate with levels of SARS-CoV-2 RNA in site-collected NP swabs [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0, 3.0 and 4.0. The sample size for phase III was also the same under protocol versions 2.0, 3.0 and 4.0 for the agents entered into the study under these protocol versions (it was originally defined in Appendix IV of protocol version 2.0 for the agent entered into the study under protocol version 2.0). The following is adapted from protocol version 4.0; further details on the assumptions and sample size calculation are provided in protocol section 10.4.

2.4.1 Phase II

For each investigational agent in phase II, the proposed sample size in 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

With respect to symptom duration, assuming 100 participants in each group will provide study diary data, the study will have 81% power to show a 33% relative reduction in median duration of symptoms, assuming:

- Log-10 Durations are normally distributed with 0.425 standard deviation;
- Wilcoxon rank sum test with two-sided 5% Type I error rate.

2.4.2 Phase III – Infused Agents

For each infused agent in phase III, the proposed sample size is 842 participants consisting of 421 participants who receive the active agent and 421 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size has been chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring during each phase of the study; additional details on monitoring can be found in protocol section 10.5. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to investigational agent or placebo or if two participants experience a Grade 4 AE deemed related to investigational

agent or placebo, enrollment to the investigational agent or placebo group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins. For infused agents, if there are no safety concerns, largely based on differences in the frequency of Grade 3 or 4 AEs between participants receiving the investigational agent and placebo, then the DSMB may recommend continuing enrollment of participants into phase III once phase II enrollment is complete. If enrollment continues to phase III, monthly (or as recommended by the DSMB) safety reviews will continue until the phase II interim efficacy analyses occur.

For infused agents, an early interim efficacy analysis will be undertaken when approximately 50% of participants (i.e. 110 of the 220 for a given investigational agent group, or 55 on active and 55 on pooled placebo) have viral shedding data in NP swabs through day 7. This review will include analyses of interim safety and will evaluate the activity of the investigational agent via assessment of graduation criteria; see section 5.4.1 for details on graduation rules.

At this early review, if graduation criteria for viral shedding at day 3 and/or day 7 are met, and/or graduation criteria for hospitalization/death based on all available data at the time of the analysis are met, then enrollment into phase III will continue pending the results of the day 28 graduation analysis that includes data from all 220 phase II participants. Otherwise, if graduation criteria are not met at this early review, enrollment to the agent will pause after phase II is fully enrolled (if the early interim analysis occurs before phase II is fully enrolled), or will pause as soon as possible (if phase III enrollment had already begun on the basis of safety data) while pending the 28 graduation analysis.

For infused agents, the DSMB will also review results from complete phase II follow-up through day 28 for all phase II participants (n=220). If these results indicate the graduation criteria have been met and there are no safety, resistance, or other concerns, then the DSMB may recommend continuation of the study for the full phase III period of evaluation.

2.5.2 Phase III - Infused Agents

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review

summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent). By-stratum summaries will also be reviewed.

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo/control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

For infused agents, because phase III enrollment may be allowed to proceed pending phase II efficacy results, it is recognized that if enrollment is fast then the analyses of phase II virology and symptom efficacy data may not be completed until after one or more of the phase III interim analyses have been undertaken. If this occurs, it is intended that the phase III stopping guidelines for efficacy and futility take precedence over enrollment pause/no pause and graduation criteria based on these analyses of phase II virology and symptom data. For example, if phase III criteria for futility are met but phase II virology efficacy data suggest that enrollment continue without pause, then the phase III criteria for futility take precedence and the DSMB may recommend termination of enrollment into the study.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death, or for differences in the primary outcome within strata. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo/control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

Additional details on interim monitoring are provided in protocol section 10.5.

2.6 Graduation to Phase III – Infused Agents

During the phase II period of the study, the DSMB will review interim safety and efficacy data to provide recommendations to the TOC via NIAID as to whether an infused investigational agent should graduate to phase III. The TOC will review DSMB recommendations, and may consider other secondary outcomes (e.g. dynamics of virologic measures and symptoms over time, or any evidence of viral rebound) in the decision to graduate an investigational agent from phase II to phase III.

The TOC will also consult with the company that owns the investigational agent, to determine the graduation decision. An independent, unblinded, group from the company will receive and review day 28 analysis data from the phase II comparisons of the investigational agent. The independent group will assist the company in deciding if the investigational agent should graduate to phase III and/or chose the dose of the phase III investigational agent. Based on these discussions and in consultation with the company, the TOC will decide whether an investigational agent enters into phase III.

3 Outcome Measures

All outcome measures are copied from the protocol version 4.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
   [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the first of two consecutive days when any symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, , and any symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, and 28.
   [For Primary Objective 3 and Secondary Objective 5]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
[Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[For Secondary Objective 7]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary on two consecutive days through day 28.
[Supportive of both Primary Objective 2 and Secondary Objective 7]

6) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or placebo, hospitalization, and death. [For Secondary Objective 1].

Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

7) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or placebo. [For Secondary Objective 2]

8) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

9) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

10) Phase II only: Level (quantitative) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, and 28.
[For Secondary Objective 5]

11) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from site-collected NP swabs at days 0, 3, 7, 14, and 28. [Supportive of both Primary Objective 3 and Secondary Objective 5]

12) Phase II and III: Level of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 1-14 and 28 in phase II, and at entry and days 3, 7, 14, and 28 in phase III.

13) Phase II and III: Quantification ((<LLoQ versus ≥LLoQ) of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [Supportive of Secondary Objective 3]

    Swabs collected at entry and days 1-14 and 28 in phase II, and at entry and days 3, 7, 10, 14, and 28 in phase III.

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from participant-collected nasal swabs daily at days 0-14 and at day 28. [Supportive of Secondary Objective 3]

Efficacy

15) Phase II only: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death from any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

17) Phase II and III: Death from any cause or hospitalization during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[For Secondary Objective 8, with follow-up beyond day 28]

Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death from any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Secondary Objective 8, with follow-up beyond day 28]

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or placebo. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or placebo. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28.
   [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
    [For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

13) Phase II only: Quantification ((<LLoQ versus ≥LLoQ) of SARS-CoV-2 RNA in blood at day 7. [Supportive of Primary Objective 3]

14) Phase II only: Level of SARS-CoV-2 RNA in blood at day 7
    [Supportive of Primary Objective 3]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to a placebo (the placebo of that investigational agent or the placebo of any other investigational agent).

- Treated Population: All participants in the Randomized Population who received any investigational agent/placebo (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus placebo. Exclusion of participants who are randomized, but who do not start their investigational agent/placebo should not introduce bias as the study is blinded. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to a placebo, who were also eligible to have received the investigational agent of interest. The comparison group will pool across all relevant placebo groups. For the primary analysis of a specific investigational agent, a supplemental analysis will restrict the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/placebo occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/placebo.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24)

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized and placebo-controlled; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLOQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or placebo are not followed.

Participants who have protocol violations, such as those who start investigational agent or placebo outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy

The following table summarizes the primary efficacy objective in phase III and the associated estimand. Further details are provided after the table.

<table>
<tr><td colspan="3">Phase III Primary Objective for Efficacy: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28.</td></tr>
<tr><td>Estimand description</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td colspan="2">Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td colspan="2">Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td colspan="2">Treated Population</td></tr>
<tr><td>Variable(s)</td><td colspan="2">Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td colspan="2">Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td colspan="2">Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td colspan="2">Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td colspan="2">Analysis approach</td></tr>
<tr><td>Ratio ((for investigational agent divided by placebo group)) of cumulative probability of death or hospitalization over 28 days.</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="3">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3 and protocol version 4.</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (from any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (from any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17 and 18, which address secondary objective 1.2.9 from the protocol, evaluate the proportion of participants who are hospitalized or died through week 24, and the proportion who died (from any cause) through week 24. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher's exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by

determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60 years and no comorbidities, ≥ 60 years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)
   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences

between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

Because some agents may be administered using injections or infusions and others will not be, the primary safety analysis will be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed separately as part of a supplementary analysis report (for week 24 outcomes) in the same manner as the primary safety outcomes.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment. Specifically, it is defined as the time (days) from day 0 (pre-treatment) to the first of two consecutive days when all symptoms scored as moderate or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for

this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and placebo arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, but not day of discharge). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). This assumes that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms.

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s)

imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to "absent". For this outcome, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are "absent" on both day 1 and day 2).

Sensitivity Analysis

No sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to placebo to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60 years and no comorbidities, ≥ 60 years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, 14, and 28).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined

as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or placebo) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/ deaths) in either arm, inference based on Fisher’s exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).

3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60 years and no comorbidities, ≥ 60 years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

    Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the time (days) from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;

3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the

repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's

exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60years and no comorbidities, ≥ 60years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2.2 Secondary Virology

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from anterior nasal swabs at each scheduled measurement time.

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

In addition, a joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., NP swabs and anterior nasal swabs).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

Levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs and anterior nasal swabs will be analyzed using participant-specific AUCs; this will be done separately for each specimen type. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: $\max[0, \log_{10}(\text{RNA})-\log_{10}(\text{LLoQ})]$, obtained at the scheduled measurement times between and including day 0 and day 28.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 28 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 28 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent placebo through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main

effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent placebo through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent placebo who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 24.

5.3.3 Exploratory Virology

The analysis of SARS-CoV-2 RNA in blood (plasma) will be done in the same manner as the secondary analysis of SARS-CoV-2 RNA from nasal swabs. See section 6.2.2 for details.

5.3.4 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

5.4.1 Phase II to Phase III Graduation Criteria

Each infused investigational agent considered in phase II will be evaluated for graduation to phase III. Graduation will be based on there being a desired level of evidence of an effect of an investigational agent versus placebo on one or more virologic and clinical outcome measures, as well as consideration of safety. The plan for these analyses will be provided in a separate document.

5.4.2 Phase III Statistical Considerations – Infused Agents

The DSMB will review interim data from the study including descriptive summaries of study conduct and adverse events, and efficacy analyses that contrast randomized arms. The primary outcome of death or hospitalization will be compared between groups using the statistical methods outlined in this SAP; the secondary outcome of death from any cause will be also be compared between randomized arms. Interim efficacy analyses are planned when approximately 220, 421, and 632 participants have been followed for the primary outcome assessed at day 28, or earlier if the total number of hospitalizations or deaths is higher than anticipated (See SAP Section 2.5 or Protocol Section 10.5).

At each interim review, the stopping boundary for the primary analysis will be determined based on the proportion of planned maximum information that is available at the given review. The statistical information (Fisher's Information) at a given review will be calculated using the inverse of the variance (square of standard error) obtained from Greenwood's formula as part of the primary analysis. The maximum information will be pre-determined using the following formula (Tsiatis AA. Statistics in medicine. 2006 Oct 15;25(19):3236-44):

$$MI = \left\{\frac{\left(Z_{\alpha/2}+Z_{\beta}\right)^2}{\delta_A^2}\right\} * (Inflation\ Factor) = \frac{(1.96+1.28)^2}{\left\{\ln\left(\frac{0.0.075}{0.15}\right)\right\}^2} * 1.03 = 22.5.$$

As a sensitivity analysis, we will assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event one day after the date they were lost to follow up. If the interpretation of the results from the primary analysis and this sensitivity analysis are substantially different, then considerations of the potential impact of delayed ascertainment of the primary endpoint will be considered, using an approach suggested by a DSMB statistician (personal correspondence A.A. Tsiatis).

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day of hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day of hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. **If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28**:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, but not day of discharge), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued under subsequent protocol versions. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions.

7.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 48).

8 Appendix 3: Statistical Considerations for AZD7442 IV

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 48).

9 Appendix 4: Statistical Considerations for AZD7742 IM

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 48).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

10.1.2 Exploratory Objectives

1) Phase II: To determine whether SNG001 reduces severity of cough or shortness of breath or difficulty breathing through study day 28.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II only: Area under the curve of *cough* and *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *cough* and *shortness of breath or difficulty breathing* symptoms in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

11 Appendix 6: Statistical Considerations for Camostat

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a spate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Version 4.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 5.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

April 15, 2021

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **4**

GLOSSARY OF TERMS **5**

1 INTRODUCTION **6**
1.1 Purpose **6**
1.2 Version History of this SAP **6**

2 STUDY OVERVIEW **6**
2.1 Study Design **6**
2.2 Randomization Process **7**
2.3 Study Objectives **8**
2.3.1 Primary Objectives 8
2.3.2 Secondary Objectives 9
2.3.3 Exploratory Objectives 9
2.4 Overview of Sample Size Considerations **10**
2.4.1 Phase II 10
2.4.2 Phase III – Infused Agents 11
2.5 Overview of Formal Interim Monitoring **11**
2.5.1 Phase II 12
2.5.2 Phase III - Infused Agents 12
2.6 Graduation to Phase III – Infused Agents **14**

3 OUTCOME MEASURES **15**
3.1 Primary Outcome Measures: Phase III **15**
3.2 Primary Outcome Measures: Phase II **15**
3.3 Secondary Outcome Measures **16**
3.4 Other Outcome Measures **18**

4 STATISTICAL PRINCIPLES **20**
4.1 General Considerations **20**

5 ANALYSIS APPROACHES **22**
5.1 Analyses of the Primary Objectives **23**
5.1.2 Phase III Primary Objective for Efficacy 23
5.1.2 Primary Safety (Phase II and III) 26
5.1.3 Primary Clinical Symptoms (Phase II) 27
5.1.4 Primary Virologic (Phase II) 30
5.2 Analyses of Secondary Objectives **33**
5.2.1 Secondary Clinical Symptoms 33
5.2.2 Secondary Virology 38
5.3 Exploratory Analyses **41**

5.3.1 New SARS-CoV-2 among Household Contacts 41
5.3.2 Hospitalization Course 42
5.3.3 Exploratory Virology 42
5.3.4 Resistance Mutations 42
5.4 Interim Analysis Considerations 43
5.4.1 Phase II to Phase III Graduation Criteria 43
5.4.2 Phase III Statistical Considerations – Infused Agents 43

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 45

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 48
7.1 Secondary Outcome Measures 48
7.2 Analysis Approaches 48

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 49
8.1 Secondary Outcome Measures 49
8.2 Analysis Approaches 49

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 50
9.1 Secondary Outcome Measures 50
9.2 Analysis Approaches 50

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 51
10.1 Objectives 51
10.1.1 Secondary Objectives 51
10.1.2 Exploratory Objectives 51
10.2 Outcome Measures 51
10.2.1 Secondary Outcome Measures 51
10.2.2 Other Outcome Measures 51
10.3 Analysis Approaches 51

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 52
11.1 Objectives 52
11.1.1 Secondary Objectives 52
11.1.2 Exploratory Objectives 52
11.2 Outcome Measures 52
11.2.1 Secondary Outcome Measures 52
11.3 Analysis Approaches 52

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 53

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 54
13.1 Secondary Outcome Measures 54
13.2 Analysis Approaches 54

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III investigations of ACTIV-2/A5401. This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP are being limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253. Version 3.0 of the SAP is developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and is not being used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applies to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents will be described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table. SAP version 4.0 applies to agents introduced in protocol version 2.0 and onward.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 4.0.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents. For infused agents, enrollment is restricted to participants at 'higher' risk for progression to severe COVID-19. For non-infused agents, enrollment is open to participants at both 'higher' and 'lower' risk of progression to severe COVID-19. See protocol for definition of 'higher' risk.

For infused agents, the study begins with a phase II evaluation, followed by a transition into a larger phase III evaluation of promising agents that 'graduate' from phase II. For non-infused agents, the same phase II study will be undertaken as for infused agents, however, the design of the phase III evaluation for non-infused agents will be developed in a later version of the protocol.

The trial has a randomized, blinded, controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient testing of new agents against placebo within the same trial infrastructure. The graduation criteria may be changed as new agents are included in the study and so analyses supporting the recommendation to graduate or otherwise are described in a separate analysis plan.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through at least week 24 to capture long-term safety information, hospitalizations or death. Study visits may be required beyond week 24, depending on the investigational agent.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 8 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

The randomization process was the same under protocol versions 2.0, 3.0 and 4.0, and is summarized in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of participants on a given investigational agent and on the comparison group for that agent. The comparison group for a given investigational agent includes all participants who were concurrently randomized to a placebo arm in the same study phase as the investigational agent of interest, and who were also eligible to have received that investigational agent.

To achieve having a similar number of participants on the active arm and in the pooled placebo comparison group for a given investigational agent, the randomization will occur in two steps.

The first randomization will be to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

The second randomization is to the (active) investigational agent or placebo for that agent within an Agent Group. For a given participant, the probability of assignment to the active agent or placebo in the second randomization depends on (1) the number of agents currently under

investigation that the participant was eligible to receive, and (2) the current study phase of the Agent Group that the participant was assigned to in the first randomization. For a participant who was assigned to an Agent Group under evaluation in phase 2, the randomization will occur at a ratio of n_2:1, where n_2 is the number of investigational agents the participant was eligible to receive that are currently under investigation in phase 2. Similarly, for a participant who was assigned to an Agent Group under evaluation in phase 3, the randomization will occur at a ratio of n_3:1, where n_3 is the number of investigational agents the participant was eligible to receive that are currently under investigation in phase 3. Here, n (the total number of investigational agents the participant was eligible to receive in the first randomization) is equal to the sum of n_2 and n_3 (i.e., $n=n_2+n_3$).

Both the first and second randomizations involve blocked stratified randomization. For non-infused agents, both the first and second randomizations are stratified by (1) time from symptom onset (≤ 5 days vs > 5 days), and (2) risk of progression to severe COVID-19 ('higher' vs 'lower'). For infused agents, both the first and second randomizations are only stratified by time from symptom onset (≤ 5 days vs > 5 days), as only 'higher' risk participants are eligible for infused agents. A participant is considered at 'higher' risk of progression to severe COVID-19 if they have a least one of several protocol-specified factors (see protocol for details).Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 4.0; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed in supplementary analysis plans. In addition, exploratory objectives 1, 4, and 12 will also be addressed in this SAP; however, other exploratory objectives will be addressed in subsequent analysis plans.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, 14, and 28 [Protocol Objective 1.1.3].

4) Phase III for infused agents only: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in nasal swabs [Protocol Objective 1.2.3].

4) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.4].

5) Phase II: To evaluate differences in SARS-CoV-2 RNA levels in NP swabs between the investigational agent versus placebo and among subgroups of the population and risk groups defined by age and comorbidities [Protocol Objective 1.2.5].
6) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To evaluate differences in symptom duration between the investigational agent versus placebo treatment groups among subgroups of the population, and risk groups defined by age and comorbidities [Protocol Objective 1.2.7].

8) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study week 24 [Protocol Objective 1.2.8].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes across the study population, notably sex, time from symptom onset to start of investigational agent, race/ethnicity, and risk groups defined by age and comorbidities [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course once a participant requires hospitalization [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phase II: To explore the impact of investigational agents on SARS-CoV-2 viremia, i.e., detection or level of SARS-CoV-2 RNA in the blood [Protocol Objective 1.3.11].

12) Phase II: To explore if levels of SARS-CoV-2 RNA in self-collected nasal swabs correlate with levels of SARS-CoV-2 RNA in site-collected NP swabs [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0, 3.0 and 4.0. The sample size for phase III was also the same under protocol versions 2.0, 3.0 and 4.0 for the agents entered into the study under these protocol versions (it was originally defined in Appendix IV of protocol version 2.0 for the agent entered into the study under protocol version 2.0). The following is adapted from protocol version 4.0; further details on the assumptions and sample size calculation are provided in protocol section 10.4.

2.4.1 Phase II

For each investigational agent in phase II, the proposed sample size in 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

With respect to symptom duration, assuming 100 participants in each group will provide study diary data, the study will have 81% power to show a 33% relative reduction in median duration of symptoms, assuming:

- Log-10 Durations are normally distributed with 0.425 standard deviation;
- Wilcoxon rank sum test with two-sided 5% Type I error rate.

2.4.2 Phase III – Infused Agents

For each infused agent in phase III, the proposed sample size is 842 participants consisting of 421 participants who receive the active agent and 421 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size has been chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring during each phase of the study; additional details on monitoring can be found in protocol section 10.5. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to investigational agent or placebo or if two participants experience a Grade 4 AE deemed related to investigational

agent or placebo, enrollment to the investigational agent or placebo group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins. For infused agents, if there are no safety concerns, largely based on differences in the frequency of Grade 3 or 4 AEs between participants receiving the investigational agent and placebo, then the DSMB may recommend continuing enrollment of participants into phase III once phase II enrollment is complete. If enrollment continues to phase III, monthly (or as recommended by the DSMB) safety reviews will continue until the phase II interim efficacy analyses occur.

For infused agents, an early interim efficacy analysis will be undertaken when approximately 50% of participants (i.e. 110 of the 220 for a given investigational agent group, or 55 on active and 55 on pooled placebo) have viral shedding data in NP swabs through day 7. This review will include analyses of interim safety and will evaluate the activity of the investigational agent via assessment of graduation criteria; see section 5.4.1 for details on graduation rules.

At this early review, if graduation criteria for viral shedding at day 3 and/or day 7 are met, and/or graduation criteria for hospitalization/death based on all available data at the time of the analysis are met, then enrollment into phase III will continue pending the results of the day 28 graduation analysis that includes data from all 220 phase II participants. Otherwise, if graduation criteria are not met at this early review, enrollment to the agent will pause after phase II is fully enrolled (if the early interim analysis occurs before phase II is fully enrolled), or will pause as soon as possible (if phase III enrollment had already begun on the basis of safety data) while pending the 28 graduation analysis.

For infused agents, the DSMB will also review results from complete phase II follow-up through day 28 for all phase II participants (n=220). If these results indicate the graduation criteria have been met and there are no safety, resistance, or other concerns, then the DSMB may recommend continuation of the study for the full phase III period of evaluation.

2.5.2 Phase III - Infused Agents

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review

summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent). By-stratum summaries will also be reviewed.

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo/control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

For infused agents, because phase III enrollment may be allowed to proceed pending phase II efficacy results, it is recognized that if enrollment is fast then the analyses of phase II virology and symptom efficacy data may not be completed until after one or more of the phase III interim analyses have been undertaken. If this occurs, it is intended that the phase III stopping guidelines for efficacy and futility take precedence over enrollment pause/no pause and graduation criteria based on these analyses of phase II virology and symptom data. For example, if phase III criteria for futility are met but phase II virology efficacy data suggest that enrollment continue without pause, then the phase III criteria for futility take precedence and the DSMB may recommend termination of enrollment into the study.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death, or for differences in the primary outcome within strata. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo/control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

Additional details on interim monitoring are provided in protocol section 10.5.

2.6 Graduation to Phase III – Infused Agents

During the phase II period of the study, the DSMB will review interim safety and efficacy data to provide recommendations to the TOC via NIAID as to whether an infused investigational agent should graduate to phase III. The TOC will review DSMB recommendations, and may consider other secondary outcomes (e.g. dynamics of virologic measures and symptoms over time, or any evidence of viral rebound) in the decision to graduate an investigational agent from phase II to phase III.

The TOC will also consult with the company that owns the investigational agent, to determine the graduation decision. An independent, unblinded, group from the company will receive and review day 28 analysis data from the phase II comparisons of the investigational agent. The independent group will assist the company in deciding if the investigational agent should graduate to phase III and/or chose the dose of the phase III investigational agent. Based on these discussions and in consultation with the company, the TOC will decide whether an investigational agent enters into phase III.

3 Outcome Measures

All outcome measures are copied from the protocol version 4.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
   [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the first of two consecutive days when any symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, , and any symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, and 28.
   [For Primary Objective 3 and Secondary Objective 5]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
[Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[For Secondary Objective 7]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary on two consecutive days through day 28.
[Supportive of both Primary Objective 2 and Secondary Objective 7]

6) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or placebo, hospitalization, and death. [For Secondary Objective 1].

Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

7) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or placebo. [For Secondary Objective 2]

8) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

9) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

10) Phase II only: Level (quantitative) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, and 28.
[For Secondary Objective 5]

11) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from site-collected NP swabs at days 0, 3, 7, 14, and 28. [Supportive of both Primary Objective 3 and Secondary Objective 5]

12) Phase II and III: Level of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 1-14 and 28 in phase II, and at entry and days 3, 7, 14, and 28 in phase III.

13) Phase II and III: Quantification ((<LLoQ versus ≥LLoQ) of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [Supportive of Secondary Objective 3]

    Swabs collected at entry and days 1-14 and 28 in phase II, and at entry and days 3, 7, 10, 14, and 28 in phase III.

14) Phase II and III: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from participant-collected nasal swabs daily at days 0-14 and at day 28. [Supportive of Secondary Objective 3]

Efficacy

15) Phase II only: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death from any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

17) Phase II and III: Death from any cause or hospitalization during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[For Secondary Objective 8, with follow-up beyond day 28]

Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death from any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Secondary Objective 8, with follow-up beyond day 28]

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or placebo. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or placebo. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28.
   [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
    [For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

13) Phase II only: Quantification (<LLoQ versus ≥LLoQ) of SARS-CoV-2 RNA in blood at day 7. [Supportive of Primary Objective 3]

14) Phase II only: Level of SARS-CoV-2 RNA in blood at day 7
    [Supportive of Primary Objective 3]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to a placebo (the placebo of that investigational agent or the placebo of any other investigational agent).

- Treated Population: All participants in the Randomized Population who received any investigational agent/placebo (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus placebo. Exclusion of participants who are randomized, but who do not start their investigational agent/placebo should not introduce bias as the study is blinded. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to a placebo, who were also eligible to have received the investigational agent of interest. The comparison group will pool across all relevant placebo groups. For the primary analysis of a specific investigational agent, a supplemental analysis will restrict the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/placebo occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/placebo.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24)

Week 48: Key visit for evaluating longer-term safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized and placebo-controlled; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLOQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or placebo are not followed.

Participants who have protocol violations, such as those who start investigational agent or placebo outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy

The following table summarizes the primary efficacy objective in phase III and the associated estimand. Further details are provided after the table.

<table>
<tr><td colspan="3">Phase III Primary Objective for Efficacy: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28.</td></tr>
<tr><td>Estimand description</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td colspan="2">Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td colspan="2">Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td colspan="2">Treated Population</td></tr>
<tr><td>Variable(s)</td><td colspan="2">Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td colspan="2">Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td colspan="2">Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td colspan="2">Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td colspan="2">Analysis approach</td></tr>
<tr><td>Ratio ((for investigational agent divided by placebo group)) of cumulative probability of death or hospitalization over 28 days.</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="3">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5).</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (from any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (from any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17 and 18, which address secondary objective 1.2.9 from the protocol, evaluate the proportion of participants who are hospitalized or died through week 24, and the proportion who died (from any cause) through week 24. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher's exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by

determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60 years and no comorbidities, ≥ 60 years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)
   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative

and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

Because some agents may be administered using injections or infusions and others will not be, the primary safety analysis will be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed separately as part of a supplementary analysis report (for week 24 outcomes) in the same manner as the primary safety outcomes.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the time (days) from day 0 (pre-treatment) to the first of two consecutive days when all symptoms scored as moderate or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for

this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and placebo arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully

completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to "absent". For this outcome, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are "absent" on both day 1 and day 2).

Sensitivity Analysis

No sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had

extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to placebo to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60 years and no comorbidities, ≥ 60 years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, 14, and 28).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-

value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as "0" if above LLoQ, and "1" if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or placebo) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/ deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.

- For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).

3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60 years and no comorbidities, ≥ 60 years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

    Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the time (days) from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;

3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the

repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's

exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification (<60years and no comorbidities, ≥ 60years or at least one comorbidity)
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity)
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2.2 Secondary Virology

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from anterior nasal swabs at each scheduled measurement time.

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

In addition, a joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., NP swabs and anterior nasal swabs).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

Levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs and anterior nasal swabs will be analyzed using participant-specific AUCs; this will be done separately for each specimen type. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: $\max[0, \log_{10}(\text{RNA})-\log_{10}(\text{LLoQ})]$, obtained at the scheduled measurement times between and including day 0 and day 28.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 28 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 28 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent placebo through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main

effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent placebo through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent placebo who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 24.

5.3.3 Exploratory Virology

The analysis of SARS-CoV-2 RNA in blood (plasma) will be done in the same manner as the secondary analysis of SARS-CoV-2 RNA from nasal swabs. See section 6.2.2 for details.

5.3.4 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

5.4.1 Phase II to Phase III Graduation Criteria

Each infused investigational agent considered in phase II will be evaluated for graduation to phase III. Graduation will be based on there being a desired level of evidence of an effect of an investigational agent versus placebo on one or more virologic and clinical outcome measures, as well as consideration of safety. The plan for these analyses will be provided in a separate document.

5.4.2 Phase III Statistical Considerations – Infused Agents

The DSMB will review interim data from the study including descriptive summaries of study conduct and adverse events, and efficacy analyses that contrast randomized arms. The primary outcome of death or hospitalization will be compared between groups using the statistical methods outlined in this SAP; the secondary outcome of death from any cause will be also be compared between randomized arms. Interim efficacy analyses are planned when approximately 220, 421, and 632 participants have been followed for the primary outcome assessed at day 28, or earlier if the total number of hospitalizations or deaths is higher than anticipated (See SAP Section 2.5 or Protocol Section 10.5).

At each interim review, the stopping boundary for the primary analysis will be determined based on the proportion of planned maximum information that is available at the given review. The proportion of planned maximum information is obtained by taking the ratio of the observed information divided by the planned maximum information (Tsiatis AA. Statistics in Medicine. 2006 Oct 15;25(19):3236-44). The planned maximum information was pre-determined using EAST software, taking into account both efficacy and futility boundaries. Assuming a one-sided 2.5% type-I error rate, 90% power, a first review at N=220 participants (coincident with Phase 2 graduation analysis), and then interim reviews at N=421 (50% of planned sample size) and N=632 (75% of planned sample size), and the final analysis after N=842 participants (if the study is not stopped earlier), the above-stated boundaries for efficacy and futility, and the assumed treatment effect of ln(0.5), the planned maximum information is 23.753. The observed statistical information at a given interim review is the inverse of the square of the standard error of the estimated treatment effect.

As outlined in the SAP, the analysis will be undertaken on a log scale and then transformed back to a risk ratio scale. The estimated treatment effect on the log scale is determined by calculating the difference between arms in log-transformed cumulative proportion estimated using Kaplan-Meier methods:

$$\hat{\delta} = \ln(\hat{p}_{active}) - \ln(\hat{p}_{placebeo}).$$

The standard error of the treatment effect on the log scale is determined by taking the square root of the sum of the variances of log-transformed cumulative proportion in each arm:

$$SE(\hat{\delta}) = \sqrt{var[ln(\hat{p}_{active})] + var[ln(\hat{p}_{placebo})]},$$

with the variances obtained using Greenwood's formula. The estimated risk ratio comparing active agent to placebo is then estimated by $\exp(\hat{\delta})$, with confidence interval calculated by taking the exponential of the confidence bounds for δ calculated on the log scale. A Wald test of the null hypothesis of no treatment effect can be constructed using the z-statistic equal to $\hat{\delta}/SE(\hat{\delta})$.

As noted above in the SAP, because of possible concerns about the validity of the analysis based on using Greenwood's approach for estimating the variance of the treatment effect estimator when the number of events is small, an alternative analysis using Fisher's exact test will be pursued if the observed events rates are smaller than 5 in one or both arms. If this arises, the proportion of planned maximum information at an interim analysis will be approximated by the proportion of the expected number of events under the trial design parameters (15% event rate in the control arm and 7.5% event rate in the placebo arm) that have been observed. Specifically, the expected number of events is 95 (15% of 421 plus 7.5% of 421). Thus, for example, if 10 events have been observed at the interim analysis, the proportion of maximal information will be approximated by 10/95=0.1053. The proportion information will be used to determine the efficacy and futility boundaries in EAST software, and critical nominal one-sided p-values corresponding to the critical Z-statistics for the efficacy and futility boundaries will be determined for comparison with the estimated nominal one-sided fisher's exact test p-value.

As a sensitivity analysis, we will assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event one day after the date they were lost to follow up.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. **If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28**:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, day of discharge, or day of withdrawal from the study during hospitalization). This assumes that the

censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued under subsequent protocol versions. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions.

7.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 48).

8 Appendix 3: Statistical Considerations for AZD7442 IV

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 48).

9 Appendix 4: Statistical Considerations for AZD7742 IM

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 48).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

10.1.2 Exploratory Objectives

1) Phase II: To determine whether SNG001 reduces severity of cough or shortness of breath or difficulty breathing through study day 28.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II only: Area under the curve of *cough* and *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *cough* and *shortness of breath or difficulty breathing* symptoms in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

11 Appendix 6: Statistical Considerations for Camostat

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a spate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 48).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Version 5.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 6.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

June 24, 2021

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **4**

GLOSSARY OF TERMS **6**

1 INTRODUCTION **7**
1.1 Purpose **7**
1.2 Version History of this SAP **7**

2 STUDY OVERVIEW **8**
2.1 Study Design **8**
2.2 Randomization Process **8**
2.3 Study Objectives **9**
2.3.1 Primary Objectives 9
2.3.2 Secondary Objectives 10
2.3.3 Exploratory Objectives 10
2.4 Overview of Sample Size Considerations **12**
2.4.1 Phase II 12
2.4.2 Phase III 12
2.5 Overview of Formal Interim Monitoring **13**
2.5.1 Phase II 13
2.5.2 Phase III 13
2.6 Graduation to Phase III **15**

3 OUTCOME MEASURES **15**
3.1 Primary Outcome Measures: Phase III **15**
3.2 Primary Outcome Measures: Phase II **16**
3.3 Secondary Outcome Measures **16**
3.4 Other Outcome Measures **19**

4 STATISTICAL PRINCIPLES **20**
4.1 General Considerations **20**

5 ANALYSIS APPROACHES **23**
5.1 Analyses of the Primary Objectives **24**
5.1.2 Phase III Primary Objective for Efficacy 24
5.1.2 Primary Safety (Phase II and III) 27
5.1.3 Primary Clinical Symptoms (Phase II) 28
5.1.4 Primary Virologic (Phase II) 32
5.2 Analyses of Secondary Objectives **35**
5.2.1 Secondary Clinical Symptoms 35
5.2.2 Secondary Virology 40
5.3 Exploratory Analyses **43**

5.3.1 New SARS-CoV-2 among Household Contacts 43
5.3.2 Hospitalization Course 44
5.3.3 Resistance Mutations 44
5.4 Interim Analysis Considerations 44
5.4.1 Phase III Statistical Considerations 44

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 46

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 49
7.1 Secondary Outcome Measures 49
7.2 Analysis Approaches 49

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 50
8.1 Secondary Outcome Measures 50
8.2 Analysis Approaches 50

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 51
9.1 Secondary Outcome Measures 51
9.2 Analysis Approaches 51

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 52
10.1 Objectives 52
10.1.1 Secondary Objectives 52
10.1.2 Exploratory Objectives 52
10.2 Outcome Measures 52
10.2.1 Secondary Outcome Measures 52
10.2.2 Other Outcome Measures 52
10.3 Analysis Approaches 52

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 53
11.1 Objectives 53
11.1.1 Secondary Objectives 53
11.1.2 Exploratory Objectives 53
11.2 Outcome Measures 53
11.2.1 Secondary Outcome Measures 53
11.3 Analysis Approaches 53

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 54

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 55
13.1 Secondary Outcome Measures 55
13.2 Analysis Approaches 55

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at ‘lower’ risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., ‘higher’ risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at ‘higher risk’	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III investigations of ACTIV-2/A5401. This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 states that enrollment to all agents (except BRII-196+BRII-198 which is already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents (though they may be evaluated in a phase III component of the study to be developed in a later protocol version that includes an active comparator agent). SAP version 5.0 therefore describes planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a placebo. In addition, SAP version 5.0 addresses some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 6.0.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized, blinded, controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient testing of new agents against placebo within the same trial infrastructure.

Version 6.0 of the protocol restricts new enrollment to agents in phase II to participants at lower risk of progression to severe COVID-19, regardless of the mode of administration of the agent. The current phase III evaluation is continuing as a placebo-controlled evaluation of the one agent that was previously approved for full phase III evaluation (BRII-196+BRII-198), and is continuing to enroll only participants at higher risk of progression to severe COVID-19. The design of the phase III evaluation for other agents will be developed in a subsequent version of the protocol and will include an active-controlled comparator instead of a placebo control.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through at least week 24 to capture long-term safety information, hospitalizations or death. Study visits may be required beyond week 24, depending on the investigational agent.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0) previous versions of the protocol), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of participants on a given investigational agent and on the comparison group for that agent. The comparison group for a given investigational agent includes all participants who were concurrently randomized to a placebo arm in the same study phase as the investigational agent of interest, and who were also eligible to have received that investigational agent.

To achieve having a similar number of participants on the active arm and in the pooled placebo comparison group for a given investigational agent, the randomization will occur in two steps.

The first randomization will be to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents (e.g., a 1:1 ratio for two agents,

1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

The second randomization is to the (active) investigational agent or placebo for that agent within an Agent Group. For a given participant, the probability of assignment to the active agent or placebo in the second randomization depends on (1) the number of agents currently under investigation that the participant was eligible to receive, and (2) the current study phase of the Agent Group that the participant was assigned to in the first randomization. For a participant who was assigned to an Agent Group under evaluation in phase II, the randomization will occur at a ratio of n_2:1, where n_2 is the number of investigational agents the participant was eligible to receive that are currently under investigation in phase II. Similarly, for a participant who was assigned to an Agent Group under evaluation in phase III, the randomization will occur at a ratio of n_3:1, where n_3 is the number of investigational agents the participant was eligible to receive that are currently under investigation in phase III. Here, n (the total number of investigational agents the participant was eligible to receive in the first randomization) is equal to the sum of n_2 and n_3 (i.e., $n=n_2+n_3$).

Both the first and second randomizations involve blocked stratified randomization. In protocol version 6.0, both the first and second randomizations are only stratified by time from symptom onset (≤ 5 days vs > 5 days), as only 'lower' risk participants are eligible agents in phase II and only 'higher' risk participants are eligible for the current agent in phase III. A participant is considered at 'higher' risk of progression to severe COVID-19 if they have a least one of several protocol-specified factors (see protocol for details). In previous versions of the protocol, in which both 'higher' and 'lower' risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group ('higher' vs 'lower'). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 6.0; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase 2, which will be addressed in supplementary analysis plans. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed in subsequent analysis plans.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, 14, and 28 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in nasal swabs [Protocol Objective 1.2.3].

4) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.4].

5) Phase II: To evaluate differences in SARS-CoV-2 RNA levels in NP swabs between the investigational agent versus placebo and among subgroups of the population [Protocol Objective 1.2.5].

6) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To evaluate differences in symptom duration between the investigational agent versus placebo treatment groups among subgroups of the population [Protocol Objective 1.2.7].

8) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study week 24 [Protocol Objective 1.2.8].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases III: To explore if the investigational agent changes the hospital course once a participant requires hospitalization [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 6.0. The sample size for phase III was also the same under protocol versions 2.0 to 6.0 for the agents entered into the study under these protocol versions (it was originally defined in Appendix IV of protocol version 2.0 for the agent entered into the study under protocol version 2.0). The following is adapted from protocol version 6.0; further details on the assumptions and sample size calculation are provided in protocol section 10.4.

2.4.1 Phase II

For each investigational agent in phase II, the proposed sample size in 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

With respect to symptom duration, assuming 100 participants in each group will provide study diary data, the study will have 81% power to show a 33% relative reduction in median duration of symptoms, assuming:

- Log-10 Durations are normally distributed with 0.425 standard deviation;
- Wilcoxon rank sum test with two-sided 5% Type I error rate.

2.4.2 Phase III

For the investigational agent currently in phase III evaluation (BRII-196+BRII-198), the proposed sample size is 842 participants consisting of 421 participants who receive the active agent and 421 participants who are concurrently randomized to placebo control. This sample size includes the enrollment that occurred during the phase II evaluation of this agent. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed and will be replaced.

This sample size has been chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring during each phase of the study; additional details on monitoring can be found in protocol section 10.5. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to investigational agent or placebo or if two participants experience a Grade 4 AE deemed related to investigational agent or placebo, enrollment to the investigational agent or placebo group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo/control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo/control. For the purpose of evaluating

statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

Additional details on interim monitoring are provided in protocol section 10.5.

2.6 Graduation to Phase III

For the investigational agent currently in phase III evaluation (BRII-196+BRII-198), details on criteria for graduation from phase II to phase III are described in protocol versions 2.0 and 3.0 (clinical sites were enrolling participants under both versions at the time of the graduation analysis). For agents in phase II, only the phase II evaluation will be undertaken, pending the design of a new phase III evaluation in a subsequent protocol version. Criteria for initiating phase III evaluation of an agent currently being evaluated in phase II will be defined in a future version of the protocol.

3 Outcome Measures

All outcome measures are copied from the protocol version 6.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
   [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the first of two consecutive days when any symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, , and any symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, and 28.
   [For Primary Objective 3 and Secondary Objective 5]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
   [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[For Secondary Objective 7]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary on two consecutive days through day 28.
[Supportive of both Primary Objective 2 and Secondary Objective 7]

6) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or placebo, hospitalization, and death. [For Secondary Objective 1].

   Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

7) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or placebo. [For Secondary Objective 2]

8) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

9) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

10) Phase II only: Level (quantitative) of SARS-CoV-2 RNA from site-collected NP swabs at days 3, 7, 14, and 28.
[For Secondary Objective 5]

11) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from site-collected NP swabs at days 0, 3, 7, 14, and 28. [Supportive of both Primary Objective 3 and Secondary Objective 5]

12) Phase III only: Level of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [For Secondary Objective 3]

    Swabs collected at entry and days 3, 7, 14, and 28 in phase III.

13) Phase III only: Quantification (<LLoQ versus ≥LLoQ) of SARS-CoV-2 RNA from participant-collected nasal swabs through day 28. [Supportive of Secondary Objective 3]

    Swabs collected at entry and days 3, 7, 14, and 28 in phase III.

14) Phase III only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from participant-collected nasal swabs through day 28. [Supportive of Secondary Objective 3]

Efficacy

15) Phase II only: Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death from any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

17) Phase II and III: Death from any cause or hospitalization during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death from any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Secondary Objective 8, with follow-up beyond day 28]

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or placebo. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or placebo. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or placebo. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to a placebo (the placebo of that investigational agent or the placebo of any other investigational agent).

- Treated Population: All participants in the Randomized Population who received any investigational agent/placebo (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus placebo. Exclusion of participants who are randomized, but who do not start their investigational agent/placebo should not introduce bias as the study is blinded. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to a placebo, who were also eligible to have received the investigational agent of interest. The comparison group will pool across all relevant placebo groups. For the primary analysis of a specific investigational agent, a supplemental analysis will restrict the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as ‘Day 0’ with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0):	First dose of investigational agent/placebo occurs. Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.
Day X:	Last day of investigational agent/placebo. Value of X depends on agent: see protocol appendices for details for each specific investigational agents.
Day 28:	Last day primary outcome may occur.
Week 24:	Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24)
Week 48:	Key visit for evaluating longer-term safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized and placebo-controlled; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLOQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or placebo are not followed.

Participants who have protocol violations, such as those who start investigational agent or placebo outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at “higher” risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy

The following table summarizes the primary efficacy objective in phase III and the associated estimand. Further details are provided after the table.

<table>
<tr><td colspan="3">Phase III Primary Objective for Efficacy: To determine if the investigational agent will prevent the composite endpoint of either hospitalization or death through study day 28.</td></tr>
<tr><td>Estimand description</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td colspan="2">Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td colspan="2">Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td colspan="2">Treated Population</td></tr>
<tr><td>Variable(s)</td><td colspan="2">Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td colspan="2">Death from any cause or hospitalization during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td colspan="2">Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td colspan="2">Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td colspan="2">Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="3">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5).</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (from any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (from any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17 and 18, which address secondary objective 1.2.9 from the protocol, evaluate the proportion of participants who are hospitalized or died through week 24, and the proportion who died (from any cause) through week 24. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher's exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by

determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

Because some agents may be administered using injections or infusions and others will not be, the primary safety analysis will be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed separately as part of a supplementary analysis report (for week 24 outcomes) in the same manner as the primary safety outcomes.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the time (days) from day 0 (pre-treatment) to the first of two consecutive days when all symptoms scored as moderate or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including “survival”

functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and placebo arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully

completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to "absent", and (b) four consecutive days with resolution of all targeted symptoms to "absent". For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are "absent" on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis will be repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but it this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to placebo to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, 14, and 28).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or placebo) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher’s exact test to compare arms will be adopted instead of using the log-

binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the time (days) from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;

3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the

repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's

exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2.2 Secondary Virology

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from anterior nasal swabs at each scheduled measurement time.

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

In addition, a joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time. Analysis will be conducted separately for each specimen type (i.e., NP swabs and anterior nasal swabs).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

Levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs and anterior nasal swabs will be analyzed using participant-specific AUCs; this will be done separately for each specimen type. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: $\max[0, \log_{10}(\text{RNA})-\log_{10}(\text{LLoQ})]$, obtained at the scheduled measurement times between and including day 0 and day 28.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 28 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 28 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent placebo through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent placebo through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent placebo who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 24.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

5.4.1 Phase III Statistical Considerations

The DSMB will review interim data from the study including descriptive summaries of study conduct and adverse events, and efficacy analyses that contrast randomized arms. The primary outcome of death or hospitalization will be compared between groups using the statistical methods outlined in this SAP; the secondary outcome of death from any cause will be also be compared between randomized arms. Interim efficacy analyses are planned when approximately 220, 421, and 632 participants have been followed for the primary outcome assessed at day 28, or earlier if the total number of hospitalizations or deaths is higher than anticipated (See SAP Section 2.5 or Protocol Section 10.5).

At each interim review, the stopping boundary for the primary analysis will be determined based on the proportion of planned maximum information that is available at the given review. The proportion of planned maximum information is obtained by taking the ratio of the observed information divided by the planned maximum information (Tsiatis AA. Statistics in Medicine. 2006

Oct 15;25(19):3236-44). The planned maximum information was pre-determined using EAST software, taking into account both efficacy and futility boundaries. Assuming a one-sided 2.5% type-I error rate, 90% power, a first review at N=220 participants (coincident with Phase II graduation analysis), and then interim reviews at N=421 (50% of planned sample size) and N=632 (75% of planned sample size), and the final analysis after N=842 participants (if the study is not stopped earlier), the above-stated boundaries for efficacy and futility, and the assumed treatment effect of ln(0.5), the planned maximum information is 23.753. The observed statistical information at a given interim review is the inverse of the square of the standard error of the estimated treatment effect.

As outlined in the SAP, the analysis will be undertaken on a log scale and then transformed back to a risk ratio scale. The estimated treatment effect on the log scale is determined by calculating the difference between arms in log-transformed cumulative proportion estimated using Kaplan-Meier methods:

$$\hat{\delta} = \ln(\hat{p}_{active}) - \ln(\hat{p}_{placebeo}).$$

The standard error of the treatment effect on the log scale is determined by taking the square root of the sum of the variances of log-transformed cumulative proportion in each arm:

$$SE(\hat{\delta}) = \sqrt{var[ln(\hat{p}_{active})] + var[ln(\hat{p}_{placebo})]},$$

with the variances obtained using Greenwood's formula. The estimated risk ratio comparing active agent to placebo is then estimated by $\exp(\hat{\delta})$, with confidence interval calculated by taking the exponential of the confidence bounds for δ calculated on the log scale. A Wald test of the null hypothesis of no treatment effect can be constructed using the z-statistic equal to $\hat{\delta}/SE(\hat{\delta})$.

As noted above in the SAP, because of possible concerns about the validity of the analysis based on using Greenwood's approach for estimating the variance of the treatment effect estimator when the number of events is small, an alternative analysis using Fisher's exact test will be pursued if the observed events rates are smaller than 5 in one or both arms. If this arises, the proportion of planned maximum information at an interim analysis will be approximated by the proportion of the expected number of events under the trial design parameters (15% event rate in the control arm and 7.5% event rate in the placebo arm) that have been observed. Specifically, the expected number of events is 95 (15% of 421 plus 7.5% of 421). Thus, for example, if 10 events have been observed at the interim analysis, the proportion of maximal information will be approximated by 10/95=0.1053. The proportion information will be used to determine the efficacy and futility boundaries in EAST software, and critical nominal one-sided p-values corresponding to the critical Z-statistics for the efficacy and futility boundaries will be determined for comparison with the estimated nominal one-sided fisher's exact test p-value.

As a sensitivity analysis, we will assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event one day after the date they were lost to follow up.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. **If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28**:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during

hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued under subsequent protocol versions. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions.

7.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 48).

8 Appendix 3: Statistical Considerations for AZD7442 IV

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, for phase II, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 48).

9 Appendix 4: Statistical Considerations for AZD7742 IM

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 48).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

10.1.2 Exploratory Objectives

1) Phase II: To determine whether SNG001 reduces severity of cough or shortness of breath or difficulty breathing through study day 28.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II only: Area under the curve of *cough* and *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *cough* and *shortness of breath or difficulty breathing* symptoms in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

11 Appendix 6: Statistical Considerations for Camostat

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a spate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 48. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 48 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 48).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Phase II/III Placebo-Controlled Study Components

Version 6.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 7.0 and

Letter of Amendment #1 to Protocol v7.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

September 13, 2021

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **4**

GLOSSARY OF TERMS **6**

1 INTRODUCTION **7**
1.1 Purpose **7**
1.2 Version History of this SAP **7**

2 STUDY OVERVIEW **8**
2.1 Study Design **8**
2.2 Randomization Process **9**
2.3 Study Objectives **10**
2.3.1 Primary Objectives 10
2.3.2 Secondary Objectives 10
2.3.3 Exploratory Objectives 11
2.4 Overview of Sample Size Considerations **12**
2.4.1 Phase II – Placebo-Controlled Superiority 12
2.4.2 Phase III – Placebo-Controlled Superiority 13
2.5 Overview of Formal Interim Monitoring **13**
2.5.1 Phase II – Placebo-Controlled Superiority 14
2.5.2 Phase III – Placebo-Controlled Superiority 14
2.6 Graduation to Phase III **15**

3 OUTCOME MEASURES **16**
3.1 Primary Outcome Measures: Phase III **16**
3.2 Primary Outcome Measures: Phase II **17**
3.3 Secondary Outcome Measures **17**
3.4 Other Outcome Measures **20**

4 STATISTICAL PRINCIPLES **22**
4.1 General Considerations **22**

5 ANALYSIS APPROACHES **24**
5.1 Analyses of the Primary Objectives **25**
5.1.2 Phase III Primary Objective for Efficacy 25
5.1.2 Primary Safety (Phase II and III) 28
5.1.3 Primary Clinical Symptoms (Phase II) 29
5.1.4 Primary Virologic (Phase II) 33
5.2 Analyses of Secondary Objectives **35**
5.2.1 Secondary Clinical Symptoms 35
5.2.2 Secondary Virology 41
5.3 Exploratory Analyses **44**

5.3.1 New SARS-CoV-2 among Household Contacts 44
5.3.2 Hospitalization Course 45
5.3.3 Resistance Mutations 45
5.4 Interim Analysis Considerations 45
5.4.1 Phase III Statistical Considerations – Placebo-Controlled Superiority 45

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 47

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 50
7.1 Randomization Details 50
7.2 Secondary Outcome Measures 50
7.3 Analysis Approaches 50

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 51
8.1 Secondary Outcome Measures 51
8.2 Analysis Approaches 51

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 52
9.1 Secondary Outcome Measures 52
9.2 Analysis Approaches 52

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 53
10.1 Objectives 53
10.1.1 Secondary Objectives 53
10.1.2 Exploratory Objectives 53
10.2 Outcome Measures 53
10.2.1 Secondary Outcome Measures 53
10.2.2 Other Outcome Measures 53
10.3 Analysis Approaches 53

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 55
11.1 Objectives 55
11.1.1 Secondary Objectives 55
11.1.2 Exploratory Objectives 55
11.2 Outcome Measures 55
11.2.1 Secondary Outcome Measures 55
11.3 Analysis Approaches 55

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 56

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 57
13.1 Secondary Outcome Measures 57
13.2 Analysis Approaches 57

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at 'lower' risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., 'higher' risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at 'higher risk'	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	
6.0	Update SAP to address changed implemented in protocol version 7.0 and Letter of Amendment 1. Specifically: - Updated SAP to reflect changes in study objectives, schedules for some evaluations, and associated outcome measures. - Updated SAP to reflect changes in interim monitoring, mainly related to changes in graduation criteria. - Updated this SAP to focus on the placebo-controlled phase III evaluation. (Note that a separate SAP will be prepared for the active-controlled phase III evaluation introduced in protocol version 7.0).	September 13, 2021

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III placebo-controlled investigations of ACTIV-2/A5401. This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the placebo-controlled components of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

An SAP, designed to address the active-controlled phase III non-inferiority part of the study, will be developed separately from this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 states that enrollment to all agents (except BRII-196+BRII-198 which is already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents. SAP version 5.0 therefore describes planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a

placebo. In addition, SAP version 5.0 addresses some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

Protocol version 7.0 introduced an open-label non-inferiority phase III design to compare investigational agents to an active-comparator among persons at higher risk of progression to hospitalization or death. The phase III design is considered separate from the phase II superiority evaluation of agents compared to placebo, among persons at lower risk for progression to hospitalization or death. Because phase II and phase III have separate designs and are evaluated among different populations, SAP version 6.0 focuses changes made under protocol version 7.0 (and letter of amendment #1) and addresses the placebo-controlled superiority phase II/III design; note, BRII-196+BRII-198 is the only agent enrolling in the placebo-controlled phase III design. A separate phase III non-inferiority SAP will also address phase III details from protocol version 7.0 (and letter of amendment #1); this SAP will apply to phase III analyses for all other agents that move to phase III.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 7.0.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized, blinded, controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient phase II and phase III testing of new agents within the same trial infrastructure.

Version 7.0 of the protocol provides for a blinded phase II evaluation of an investigational agent compared to placebo among participants at lower risk for progression to hospitalization or death, regardless of the mode of administration of the agent; for some agents, enrollment to higher risk participants in phase II was allowed under earlier protocol versions. Agents that graduate to phase III (after initiation of this protocol version) will be evaluated in persons at higher risk for progression to hospitalization or death for non-inferiority to an active comparator, the monoclonal antibody cocktail of casirivimab plus imdevimab (REGEN-COV, Regeneron), which has been shown to be effective in this population in preventing hospitalization or death. This version of the protocol also provides for continued follow-up of participants enrolled into a placebo-controlled phase III trial evaluating the combination monoclonal antibody agent, BRII-196 + BRII-198.

When two or more agents are being evaluated in the same phase of the study, the trial design includes sharing of the control group (placebo in phase II and active comparator in phase III) for efficient evaluation of each agent. Note: enrollment to BRII-196+BRII-198 will not coincide with enrollment to other agents in phase III.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through at least week 72 weeks in phase II and phase III.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0) previous versions of the protocol), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

The phase II trial and the phase III trial involve different populations and will have separate randomizations. However, the structure of the randomization process will be the same for each of the two trials, as described in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of currently randomized participants on a given investigational agent and on the comparator group for that agent (i.e. combining participants who were eligible to receive the agent but who were randomized to any of the available placebos in phase II or to the active comparator in phase III).

To achieve having a similar number of participants on the active arm and in the pooled comparator group for a given investigational agent, the randomization will occur in two steps within each trial.

The first randomization will be to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents in the trial (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

Immediately following the first randomization, participants are randomized within an Agent Group in the second randomization to the (active) investigational agent or appropriate comparator (the matching placebo for agents in phase II, or the active comparator for agents in phase III). For a given participant, the probability of assignment to the active agent or comparator in the second randomization depends on the number of agents currently under investigation that the participant was eligible to receive, as phase II and phase III have distinct populations (phase II is restricted to those at lower risk of progression to hospitalization and death, and phase III is restricted to those at higher risk) agent phase is accounted for in the participant eligibility. .

Both the first and second randomizations involve blocked stratified randomization. In phase II, both the first and second randomizations are stratified by time from symptom onset (≤ 5 days vs > 5 days), however, in previous versions of the protocol, in which both ‘higher’ and ‘lower’ risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group (‘higher’ vs ‘lower’). Protocol version 7.0 introduced different stratification factors for phase III, with both the randomization steps stratified by country; previous versions of the protocol included only stratification by time from symptom

onset (≤ 5 days vs > 5 days). In addition, in previous versions of the protocol, in which both 'higher' and 'lower' risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group ('higher' vs 'lower'). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 7.0 (and letter of amendment #1); corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase 2, which will be addressed in supplementary analysis plans. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed in subsequent analysis plans.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, and 14 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in NP swabs [Protocol Objective 1.2.3].

4) Phase III: To determine the efficacy of the investigational agent to increase the proportion of participants with NP SARS-CoV-2 RNA below the LLoQ at study day 3 [Protocol Objective 1.2.4]

5) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.5].

6) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72 [Protocol Objective 1.2.7].

8) Phase III: To evaluate if the investigational agent reduces the time to sustained symptom resolution through study day 28 [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of hospitalization or death through study day 28, excluding hospitalizations that are determined to be unrelated to COVID-19.

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes and differences between investigational agent and control (placebo in phase II and active comparator in phase III) across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course in those hospitalized [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 7.0. The sample size for the placebo-controlled superiority phase III design was also the same under protocol versions 2.0 to 6.0 (it was originally defined in Appendix IV of protocol version 2.0 for the agent entered into the study under protocol version 2.0) and is currently detailed in Appendix V of protocol version 7.0 for the BRII-196+BRII-198 agent. Details on the sample size for the non-inferiority phase III design are provided in the phase III non-inferiority SAP and are discussed in in protocol version 7.0 section 10.4. The following sections reflect the placebo-controlled phase II and phase III designs.

2.4.1 Phase II – Placebo-Controlled Superiority

For each investigational agent in phase II, the proposed sample size is 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

With respect to symptom duration, assuming 100 participants in each group will provide study diary data, the study will have 81% power to show a 33% relative reduction in median duration of symptoms, assuming:

- Log-10 Durations are normally distributed with 0.425 standard deviation;
- Wilcoxon rank sum test with two-sided 5% Type I error rate.

2.4.2 Phase III – Placebo-Controlled Superiority

The proposed sample size is 842 participants consisting of 421 participants who receive the active agent and 421 participants who are concurrently randomized to placebo control. This sample size includes the enrollment that occurred during the phase II evaluation of an agent. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size has been chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring of the placebo-controlled phase II and phase III design during each phase of the study; additional details on phase II monitoring can be found in protocol section 10.5, and in Appendix V for phase III monitoring. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP. Details on interim monitoring for the phase III non-inferiority design are outlined in the phase III non-inferiority SAP and in protocol version 7.0 section 10.5.

Regardless of study phase, in the event that there is any death deemed related to investigational agent or placebo or if two participants experience a Grade 4 AE deemed related to investigational agent or placebo, enrollment to the investigational agent or placebo group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II – Placebo-Controlled Superiority

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III – Placebo-Controlled Superiority

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo/control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely

then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo/control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

2.6 Graduation to Phase III

The following applies to investigational agents that have been not assessed for graduation to phase III under prior versions of the protocol (version 1.0 to 6.0); note BRII-196+BRII-198 was previously assessed for graduation to phase III under protocol versions 2.0 and 3.0 (clinical sites were enrolling participants under both versions at the time of the graduation analysis).

Each investigational agent that is being considered for evaluation in phase III will be evaluated for safety, for activity in reducing COVID-19 symptoms and hospitalization/death, and for activity in reducing SARS-CoV-2 RNA shedding. An analysis to determine if an agent should graduate from

phase II, and enter phase III, will be conducted when 220 participants assigned to the agent or concurrent placebo in phase II evaluation have completed their Day 7 evaluations and have the required data available in the database. Additional interim graduations may be assessed for some agents, see agent-specific appendices in protocol version 7.0 for details.

The DSMB will review unblinded data and make recommendations to NIAID (as trial sponsor) and to the TOC, indicating whether graduation criteria have been met. The recommendation for an agent to enter phase III evaluation will be made by the TOC in discussion with the collaborating company; the collaborating company that is responsible for the agent will decide whether or not to adopt the recommendation. The TOC and collaborating company will also consider which dose to recommend for evaluation in phase III, for investigational agents with more than one dose under evaluation in phase II. NIAID/DAIDS, as the sponsor of the study, will make the final determination regarding graduation of the study product.

The TOC may recommend an agent move directly into phase III, without evaluation in phase II in ACTIV-2, if there is sufficient safety and efficacy data supporting phase III evaluation available from outside of the trial. These agents will not undergo graduation analyses.

Graduation criteria and statistical considerations are discussed in the Graduation Rules SAP.

3 Outcome Measures

All outcome measures are copied from the protocol version 7.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo. [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the number of days from start of investigational treatment to the first of two consecutive days when any symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, and any symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: At each of days 3, 7 and 14 quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs.
   [For Primary Objective 3]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
   [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2]

Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2 and for Secondary Objective 8]

Duration defined as the number of days from start of investigational treatment to the first of four consecutive days when all symptoms are scored as absent. Targeted symptoms as defined in the primary phase II clinical (symptom duration) outcome.

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of two consecutive days that a participant reported return to usual (pre-COVID) health.

7) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

8) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or placebo, hospitalization, and death. [For Secondary Objective 1].

Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first

day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

9) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or placebo. [For Secondary Objective 2]

10) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

11) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

12) Phase III only: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs at day 3. [Support of Primary Objective 3]

13) Phase II and III: Level (quantitative) of SARS-CoV-2 RNA from staff-collected NP swabs at days 3, 7, and 14 in phase II and at day 3 in phase III.8.
[For Secondary Objective 3]

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from staff-collected NP swabs at days 0, 3, 7, and 14. [Supportive of both Primary Objective 3 and Secondary Objective 3]

Efficacy

15) Phase II only: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.
[Supportive of Primary Objective 4]

17) Phase II and III: Death due to any cause or hospitalization due to any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo.
[For Secondary Objective 8, with follow-up beyond day 28]

Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death due to any cause or hospitalization due to any cause during the 72-week period from and including the day of the first dose of investigational agent or placebo. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

19) Phase II and III: Death due to any cause during the 24-week period from and including the day of the first dose of investigational agent or placebo.
    [Supportive of Secondary Objective 8, with follow-up beyond day 28]

20) Phase II and III: Death due to any cause during the 72-week period from and including the day of the first dose of investigational agent or placebo.
    [Supportive of Secondary Objective 8, with follow-up beyond day 28]

21) Phase II and III: Death due to any cause or hospitalization due to any cause, excluding hospitalizations that are deemed unrelated to COVID-19, during the 28-day period from and including the day of the first dose of investigational agent or placebo.
    [For Secondary Objective 9]

    Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or placebo. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or placebo.
   [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or placebo.
   [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or placebo.
   [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

Ordinal scale defined as:
Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to a placebo (the placebo of that investigational agent or the placebo of any other investigational agent).

- Treated Population: All participants in the Randomized Population who received any investigational agent/placebo (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus placebo. Exclusion of participants who are randomized, but who do not start their investigational agent/placebo should not introduce bias as the study is blinded. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to a placebo, who were also eligible to have received the investigational agent of interest. The comparison group will pool across all relevant placebo groups. For the primary analysis of a specific investigational agent, a supplemental analysis will restrict the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/placebo occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/placebo.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24).

Week 72: Key visit for evaluation longer-term efficacy and safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized and placebo-controlled; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLOQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or placebo are not followed.

Participants who have protocol violations, such as those who start investigational agent or placebo outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at "higher" risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the placebo-controlled superiority design. Further details are provided after the table.

<table>
<tr><td colspan="3">Phase III Primary Objective for Efficacy: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.</td></tr>
<tr><td>Estimand description</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td colspan="2">Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td colspan="2">Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td colspan="2">Treated Population</td></tr>
<tr><td>Variable(s)</td><td colspan="2">Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td colspan="2">Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td colspan="2">Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td colspan="2">Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td colspan="2">Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="3">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5).</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (due to any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17,18, 19, 20 and 21, , evaluate the proportion of participants who are hospitalized or died through week 24, the proportion who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days) or through 504 days (i.e. 72 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher’s exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis

will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences

between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

Because some agents may be administered using injections or infusions and others will not be, the primary safety analysis will be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed separately as part of a supplementary analysis report (for week 24 outcomes) in the same manner as the primary safety outcomes.

Additional longer-term safety outcomes may be assessed, see agent-specific appendices for details.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the number of days from start of investigational agent (day 0, pre-treatment) to the first of two consecutive days when all symptoms scored as moderate or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure

criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and placebo arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is

available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to "absent", and (b) four consecutive days with resolution of all targeted symptoms to "absent" (i.e., secondary outcome measure 5). For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are "absent" on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis will be repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but it this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to placebo to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, and 14).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or placebo) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher’s exact test to compare arms will be adopted instead of using the log-

binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
8) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the number of days from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the

number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;
3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study

day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

Duration of Symptoms

In support of the symptom duration outcome in phase III, the analysis will be repeated using the same approach described in the primary symptom duration analysis for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to "absent." To address secondary objective 8, and in supportive of the symptom duration outcome in phase III, a similar TTE outcome measure will also be examined defined as time to four consecutive days with resolution of all targeted symptoms to "absent," (i.e. secondary outcome measure 5).

Return to Usual Health

The analysis of return to usual health will be repeated using the same approach described above for a similar TTE outcome measures defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.2.2 Secondary Virology

The schedule of evaluations in protocol version 7 indicates that only NP swabs will collected in both phase II and phase III, and therefore only analyses of SARS-CoV-2 RNA from NP swabs are outlined below. Some agents may have completed enrollment in phase II prior to implementing protocol version 7.0, and therefore may have additional specimens collected for SARS-CoV-2 RNA testing. If analyses of these additional specimens are pursued, then the approach will be as defined in the previous version of the SAP.

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent placebo, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or placebo are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from staff-collected NP swabs at entry and day 3.

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ).

In addition, a joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time for staff-collected NP swabs.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of

detection but below the LLoQ will be imputed as the second lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

In phase II only, levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs will be analyzed using participant-specific AUCs. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: $\max[0, \log_{10}(\text{RNA})-\log_{10}(\text{LLoQ})]$, obtained at the scheduled measurement times between and including day 0 and day 28.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 28 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 28 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

<u>Sensitivity Analyses</u>

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent placebo through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent placebo through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent placebo who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 24.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

5.4.1 Phase III Statistical Considerations – Placebo-Controlled Superiority

The DSMB will review interim data from the study including descriptive summaries of study conduct and adverse events, and efficacy analyses that contrast randomized arms. The primary outcome of death or hospitalization will be compared between groups using the statistical methods outlined in this SAP; the secondary outcome of death from any cause will be also be compared between randomized arms. Interim efficacy analyses are planned when approximately 220, 421, and 632 participants have been followed for the primary outcome assessed at day 28, or earlier if the total number of hospitalizations or deaths is higher than anticipated (See SAP Section 2.5 or Protocol Section 10.5).

At each interim review, the stopping boundary for the primary analysis will be determined based on the proportion of planned maximum information that is available at the given review. The proportion of planned maximum information is obtained by taking the ratio of the observed information divided by the planned maximum information (Tsiatis AA. Statistics in Medicine. 2006

Oct 15;25(19):3236-44). The planned maximum information was pre-determined using EAST software, taking into account both efficacy and futility boundaries. Assuming a one-sided 2.5% type-I error rate, 90% power, a first review at N=220 participants (coincident with Phase II graduation analysis), and then interim reviews at N=421 (50% of planned sample size) and N=632 (75% of planned sample size), and the final analysis after N=842 participants (if the study is not stopped earlier), the above-stated boundaries for efficacy and futility, and the assumed treatment effect of ln(0.5), the planned maximum information is 23.753. The observed statistical information at a given interim review is the inverse of the square of the standard error of the estimated treatment effect.

As outlined in the SAP, the analysis will be undertaken on a log scale and then transformed back to a risk ratio scale. The estimated treatment effect on the log scale is determined by calculating the difference between arms in log-transformed cumulative proportion estimated using Kaplan-Meier methods:

$$\hat{\delta} = \ln(\hat{p}_{active}) - \ln(\hat{p}_{placebeo}).$$

The standard error of the treatment effect on the log scale is determined by taking the square root of the sum of the variances of log-transformed cumulative proportion in each arm:

$$SE(\hat{\delta}) = \sqrt{var[ln(\hat{p}_{active})] + var[ln(\hat{p}_{placebo})]},$$

with the variances obtained using Greenwood's formula. The estimated risk ratio comparing active agent to placebo is then estimated by $\exp(\hat{\delta})$, with confidence interval calculated by taking the exponential of the confidence bounds for δ calculated on the log scale. A Wald test of the null hypothesis of no treatment effect can be constructed using the z-statistic equal to $\hat{\delta}/SE(\hat{\delta})$.

As noted above in the SAP, because of possible concerns about the validity of the analysis based on using Greenwood's approach for estimating the variance of the treatment effect estimator when the number of events is small, an alternative analysis using Fisher's exact test will be pursued if the observed events rates are smaller than 5 in one or both arms. If this arises, the proportion of planned maximum information at an interim analysis will be approximated by the proportion of the expected number of events under the trial design parameters (15% event rate in the control arm and 7.5% event rate in the placebo arm) that have been observed. Specifically, the expected number of events is 95 (15% of 421 plus 7.5% of 421). Thus, for example, if 10 events have been observed at the interim analysis, the proportion of maximal information will be approximated by 10/95=0.1053. The proportion information will be used to determine the efficacy and futility boundaries in EAST software, and critical nominal one-sided p-values corresponding to the critical Z-statistics for the efficacy and futility boundaries will be determined for comparison with the estimated nominal one-sided fisher's exact test p-value.

As a sensitivity analysis, we will assume all participants who prematurely discontinue the study prior to day 28, who are unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event one day after the date they were lost to follow up.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. **If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28**:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during

hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued under subsequent protocol versions. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions.

7.1 Randomization Details

Phase III is stratified by time from symptom onset (≤ 5 days versus > 5 days).

7.2 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.3 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 72).

8 Appendix 3: Statistical Considerations for AZD7442 IV

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 72).

9 Appendix 4: Statistical Considerations for AZD7742 IM

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 72).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

10.1.2 Exploratory Objectives

1) Phase II and III: To determine whether SNG001 reduces severity of cough or shortness of breath or difficulty breathing through study day 28.

2) Phase II and III: To determine whether SNG001 reduces a COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 among individuals who report moderate to severe shortness of breath or difficulty breathing at day 0.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II and III: Area under the curve of *cough* and *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *cough* and *shortness of breath or difficulty breathing* symptoms in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e.

will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

To address SNG001-specific exploratory objective 2, the COVID-19 severity ranking outcome will compared between arms using the same methods outlined for the secondary analysis of this outcome measure, but restricted to be among those with moderate to severe shortness of breath or difficult breathing at day 0,

11 Appendix 6: Statistical Considerations for Camostat

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a separate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 72).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Phase II/III Placebo-Controlled and Phase III Active-Controlled

Study Components

Version 7.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 7.0 and

Letter of Amendment #1 to Protocol v7.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

October 24, 2021

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **4**

GLOSSARY OF TERMS **6**

1 INTRODUCTION **7**
1.1 Purpose **7**
1.2 Version History of this SAP **7**

2 STUDY OVERVIEW **8**
2.1 Study Design **8**
2.2 Randomization Process **9**
2.3 Study Objectives **10**
2.3.1 Primary Objectives 10
2.3.2 Secondary Objectives 10
2.3.3 Exploratory Objectives 11
2.4 Overview of Sample Size Considerations **12**
2.4.1 Phase II – Placebo-Controlled Superiority 12
2.4.2 Phase III – Placebo-Controlled Superiority Trial 13
2.4.3 Phase III – Active-Controlled Non-Inferiority Trial 13
2.5 Overview of Formal Interim Monitoring **16**
2.5.1 Phase II – Placebo-Controlled Superiority 17
2.5.2 Phase III – Placebo-Controlled Superiority 17
2.5.3 Phase III – Active-Controlled Non-Inferiority 18
2.6 Graduation to Phase III **20**

3 OUTCOME MEASURES **21**
3.1 Primary Outcome Measures: Phase III **21**
3.2 Primary Outcome Measures: Phase II **21**
3.3 Secondary Outcome Measures **22**
3.4 Other Outcome Measures **25**

4 STATISTICAL PRINCIPLES **26**
4.1 General Considerations **26**

5 ANALYSIS APPROACHES **29**
5.1 Analyses of the Primary Objectives **29**
5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation 29
5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation 34
5.1.2 Primary Safety (Phase II and III) 36
5.1.3 Primary Clinical Symptoms (Phase II) 37

5.1.4 Primary Virologic (Phase II) 41
5.2 Analyses of Secondary Objectives 43
5.2.1 Secondary Clinical Symptoms 43
5.2.2 Secondary Virology 49
5.3 Exploratory Analyses 52
5.3.1 New SARS-CoV-2 among Household Contacts 52
5.3.2 Hospitalization Course 52
5.3.3 Resistance Mutations 53
5.4 Interim Analysis Considerations 53

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 54

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 57
7.1 Randomization Details 57
7.2 Secondary Outcome Measures 57
7.3 Analysis Approaches 57

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 59
8.1 Secondary Outcome Measures 59
8.2 Analysis Approaches 59

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 60
9.1 Secondary Outcome Measures 60
9.2 Analysis Approaches 60

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 61
10.1 Objectives 61
10.1.1 Secondary Objectives 61
10.1.2 Exploratory Objectives 61
10.2 Outcome Measures 61
10.2.1 Secondary Outcome Measures 61
10.2.2 Other Outcome Measures 61
10.3 Analysis Approaches 61

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 63
11.1 Objectives 63
11.1.1 Secondary Objectives 63
11.1.2 Exploratory Objectives 63
11.2 Outcome Measures 63
11.2.1 Secondary Outcome Measures 63
11.3 Analysis Approaches 63

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 64

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 65
13.1 Secondary Outcome Measures 65
13.2 Analysis Approaches 65

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at ‘lower’ risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., ‘higher’ risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at ‘higher risk’	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	
6.0	Update SAP to address changed implemented in protocol version 7.0 and Letter of Amendment 1. Specifically: - Updated SAP to reflect changes in study objectives, schedules for some evaluations, and associated outcome measures. - Updated SAP to reflect changes in interim monitoring, mainly related to changes in graduation criteria. - Updated this SAP to focus on the placebo-controlled phase III evaluation. (Note that a separate *revision to the* SAP will be prepared for the active-controlled phase III evaluation introduced in protocol version 7.0). [Note that the italicized text was added in version 7.0 of the SAP].	September 13, 2021
7.0	Updated SAP with the following major changes: - To add details that are specific to the active-controlled phase 3 evaluation of agents based on protocol version 7.0 and Letter of Amendment 1. - To edit some text to provide clarity concerning the analysis approaches which are the same regardless of whether a placebo control or active control is involved. In part, to achieve this, the terminology "comparator intervention" is often used. - To replace the previous section 5.4 concerning interim analysis considerations for the placebo-controlled phase III trial (which have been completed) with interim analysis considerations for the active-controlled phase III trial. - To indicate exclusion from analysis of viral shedding results from samples labelled as "Thawed', 'Destroyed', 'Quantity Not Sufficient' or "Invalid Specimen' as approved by the trial sponsor. - To focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.	October 24, 2021

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III placebo-controlled investigations of ACTIV-2/A5401, as well as the phase III active-controlled investigation introduced in protocol version 7.0. This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 stated that enrollment to all agents (except BRII-196+BRII-198 which was already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents. SAP version 5.0 therefore described planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a placebo. In addition, SAP version 5.0 addressed some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

Protocol version 7.0 introduced an open-label non-inferiority phase III design to compare investigational agents to an active-comparator among persons at higher risk of progression to hospitalization or death. This phase III evaluation is separate from the phase II superiority evaluation of agents compared to placebo among persons at lower risk for progression to hospitalization or death. Changes introduced in SAP version 6.0 focused on changes made under protocol version 7.0 (and letter of amendment #1) that related to the placebo-controlled superiority phase II/III design (note: BRII-196+BRII-198 is the only agent that enrolled in the placebo-controlled phase III design). Changes introduced in SAP version 7.0 address the introduction of the active-controlled non-inferiority phase III trial in protocol version 7.0 (and letter of amendment #1). SAP version 7.0 also introduces the exclusion from statistical analysis of results generated from problematic virologic samples based on a decision made by the DAIDS and study team. In addition, section 5.4 concerning interim analysis considerations was revised to replace considerations for the placebo-controlled phase III trial for which DSMB monitoring has been completed with considerations for the active-controlled phase III trial. Finally, adjustments were made to focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 7.0 and letter of amendment 1.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient phase II and phase III testing of new agents within the same trial infrastructure.

Version 7.0 of the protocol provides for a blinded phase II evaluation of an investigational agent compared to placebo among participants at lower risk for progression to hospitalization or death, regardless of the mode of administration of the agent; for some agents, enrollment to higher risk participants in phase II was allowed under earlier protocol versions. Agents that graduate to phase III (after initiation of this protocol version) will be evaluated in persons at higher risk for progression to hospitalization or death for non-inferiority to an active comparator, the monoclonal antibody combination of casirivimab plus imdevimab (REGEN-COV, Regeneron), which has been shown to be effective in this population in preventing hospitalization or death. Protocol version 7.0 also provides for continued follow-up of participants enrolled into a placebo-controlled phase III trial evaluating the combination monoclonal antibody agent, BRII-196 + BRII-198.

When two or more agents are being evaluated in the same phase of the study, the trial design includes sharing of the control group (placebo in phase II and active comparator in phase III) for

efficient evaluation of each agent. Note that enrollment to BRII-196+BRII-198 did not coincide with enrollment to other agents in phase III.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through at least week 72 weeks in phase II and phase III.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0) previous versions of the protocol), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

The phase II trial and the active-controlled phase III trial involve different populations and have separate randomizations. However, the structure of the randomization process is the same for each of the two trials, as described in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of concurrently randomized participants on a given investigational agent and on the comparator group for that agent (i.e. combining participants who were eligible to receive the agent but who were randomized to any of the available placebos in phase II or to the active comparator in phase III).

To achieve having a similar number of participants on the active arm and in the pooled comparator group for a given investigational agent, the randomization occurs in two steps within each trial.

The first randomization is to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents in the trial (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). Trial phase for an agent is accounted for in the participant eligibility (i.e. by the classification of their risk for hospitalization/death as lower or higher). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

Immediately following the first randomization, participants are randomized within an Agent Group in the second randomization to the (active) investigational agent or appropriate comparator (the matching placebo for agents in phase II, or the active comparator for agents in phase III). For a given participant, the probability of assignment to the active agent or comparator in the second randomization depends on the number of agents currently under investigation that the participant was eligible to receive, as phase II and phase III have distinct populations (phase II is restricted to those at lower risk of progression to hospitalization and death, and phase III is restricted to those at higher risk).

Both the first and second randomizations involve blocked stratified randomization. In phase II, both the first and second randomizations are stratified by time from symptom onset (≤5 days vs >5 days), however, in previous versions of the protocol, in which both 'higher' and 'lower' risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group ('higher' vs 'lower'). In the active-controlled phase III trial introduced in protocol version 7.0, both randomization steps are stratified by country. Under previous versions of the protocol for the placebo-controlled phase III, both randomization steps were only stratified by time from symptom onset (≤ 5 days vs > 5 days). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 7.0 (and letter of amendment #1); corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase 2, which will be addressed outside of this SAP. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed separately.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, and 14 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in NP swabs [Protocol Objective 1.2.3].

4) Phase III: To determine the efficacy of the investigational agent to increase the proportion of participants with NP SARS-CoV-2 RNA below the LLoQ at study day 3 [Protocol Objective 1.2.4]

5) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.5].

6) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72 [Protocol Objective 1.2.7].

8) Phase III: To evaluate if the investigational agent reduces the time to sustained symptom resolution through study day 28 [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of hospitalization or death through study day 28, excluding hospitalizations that are determined to be unrelated to COVID-19 [Protocol Objective 1.2.9 (introduced in letter of amendment 1 to protocol version 7.0)].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes and differences between investigational agent and control (placebo in phase II and active comparator in phase III) across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course in those hospitalized [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 7.0. The sample size for the placebo-controlled superiority phase III design was also the same under protocol versions 2.0 to 6.0 (it was originally defined in Appendix IV of protocol version 2.0 for the agent entered into the study under protocol version 2.0) and is currently detailed in Appendix V of protocol version 7.0 for the BRII-196+BRII-198 agent. Details on the sample size for the non-inferiority active-controlled phase III design are from protocol version 7.0 section 10.4.

2.4.1 Phase II – Placebo-Controlled Superiority

For each investigational agent in phase II, the proposed sample size is 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

2.4.2 Phase III – Placebo-Controlled Superiority Trial

The proposed sample size is 842 participants consisting of 421 participants who receive the active agent and 421 participants who are concurrently randomized to placebo control. This sample size includes the enrollment that occurred during the phase II placebo-controlled evaluation of an agent. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size has been chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.4.3 Phase III – Active-Controlled Non-Inferiority Trial

The active-controlled Phase III trial is focused on a non-inferiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus an active comparator agent, specifically the monoclonal antibody combination of casirivimab plus imdevimab. The non-inferiority margin for the absolute difference in proportion hospitalized/dead is 3% (investigational agent minus active comparator agent); the rationale for this choice is described in Section 3.1 of protocol version 7.0. Non-inferiority will be considered established if a two-sided exact 95% confidence interval for the absolute difference is entirely below 3%. Details of the construction of the confidence interval are in section 10.6 pf the protocol and are included further below in this SAP.

The sample size differs between infused investigational agents (600 for the investigational agent and 600 for the concurrently randomized active comparator) and non-infused investigational agents (800 per arm instead of 600 per arm). The rationale for this is that there may be broader clinical utility for non-infused agents such that a slightly higher true hospitalization/death rate may be tolerated in clinical practice.

Sample Size Justification for Infused Investigational Agents

For the evaluation of a specific infused investigational agent, the sample size is 1200 participants including approximately 600 participants randomized to receive the infused investigational agent and approximately 600 participants (who were eligible to receive the infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size has been chosen to provide close to 90% power to establish non-inferiority assuming that the true proportion hospitalized/dead for both the infused investigational agent and the active comparator agent is 2.3%. The rate of 2.3% is based on the observed proportion for casirivimab plus imdevimab combining across doses in the subpopulation of the Regeneron COV-2067 clinical trial who met the criteria for being at high risk of progression to hospitalization/death (FDA communication to DAIDS/NIAID, May 2021). No adjustment for loss to follow-up is made in the sample size as the primary analysis will be based on the observed number of hospitalizations divided by the number of participants who initiated study treatment. In addition, the impact of any loss to follow-up is expected to be minimal as there will be regular contact between research site staff and participants (or their secondary contacts) and previous experience in the study and other trials has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up).

The potential power of the study was evaluated in two ways using the PASS version 15 sample size calculation software. Both used a non-inferiority hypothesis testing approach based on use of the Miettinen and Nurminen score test statistic (which is the basis for calculating the confidence interval used for analysis in this study). The first ignored interim monitoring but used a binomial enumeration method to calculate power and type I error rates. Use of the binomial enumeration method takes account of the discreteness of the binomial distribution (rather than using a normal approximation to the binomial distribution) which may be important in the setting of low hospitalization/death probabilities. Using this approach gave a power of 90.2%. The second approach did not use a binomial enumeration but took account of interim analyses using a standard implementation of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 90.0% (width of 95% confidence interval around this simulation-based value was 0.12%). Based on these two approaches, it is anticipated that the study will have close to 90% power to show non-inferiority for an infused investigational agent assuming that it truly has the same 2.3% hospitalization/death rate as the active comparator agent.

The PASS software was also used to illustrate how the power of the study might change for various scenarios which differ from the scenario assumed (see Table 2.4.3-1). This was undertaken using the first of the two approaches for evaluating prior mentioned above (i.e., using the binomial enumeration approach). Looking at the top part of the table in which both the infused investigational agent and the active comparator agent have the same underlying true hospitalization/ death rate, the power is decreased if the true rate is above the assumed 2.3%, but increased if the true rate is less than 2.3%. If the true rate is 3%, then the power is still above 80%, but if the true rate is 4% it is reduced to 73%.

The middle and lower parts of the table show scenarios in which the infused investigational agent has a true hospitalization/death rate of 0.5% or 1% worse than the active comparator agent,

respectively. If the true rate for the active comparator agent is 2.3% and is 2.8% for the infused investigational agent (i.e., 0.5% worse), then the power is reduced to 73%. If the true rate for the active comparator agent is 2.3% and is 3.3% for the infused investigational agent (i.e., 1% worse), then the power is reduced to 50%

Table 2.4.3-1: Power for various scenarios based on non-inferiority hypothesis testing using the likelihood score test statistic (Miettinen and Nurminen method) with binomial enumeration of power and Type I error rate. All scenarios use a 3% non-inferiority margin and one-sided Type-I error rate of 0.025 with a sample size of 600 participants receiving an infused investigational agent and 600 participants receiving the active comparator agent. Power in practice will be slightly reduced from the values shown due to interim monitoring.

Same Underlying True Hospitalization/Death Rate For Active Comparator Agent and Infused Investigational Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
99.4%	1%	1%	2.2%
97.3%	1.5%	1.5%	2.2%
93.2%	2%	2%	2.3%
90.2%	2.3%	2.3%	2.4%
88.1%	2.5%	2.5%	2.4%
83.1%	3%	3%	2.4%
78.1%	3.5%	3.5%	2.4%
73.2%	4%	4%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 0.5% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
92.5%	1%	1.5%	2.2%
85.2%	1.5%	2%	2.2%
77.4%	2%	2.5%	2.3%
73.1%	2.3%	2.8%	2.4%
70.5%	2.5%	3%	2.4%
64.8%	3%	3.5%	2.4%
59.6%	3.5%	4%	2.4%
55.0%	4%	4.5%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 1% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
71.3%	1%	2%	2.2%
61.7%	1.5%	2.5%	2.2%
54.0%	2%	3%	2.3%
50.4%	2.3%	3.3%	2.4%
48.4%	2.5%	3.5%	2.4%

Same Underlying True Hospitalization/Death Rate For Active Comparator Agent and Infused Investigational Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
44.0%	3%	4%	2.4%
40.0%	3.5%	4.5%	2.4%
36.7%	4%	5%	2.4%
*Actual type I error rate is slightly lower than assumed rate of 2.5% because of discreteness of the binomial distribution.			

Sample Size Justification for Non-infused Investigational Agents

For the evaluation of a specific non-infused investigational agent, the sample size will include approximately 800 participants randomized to receive the non-infused investigational agent and approximately 800 participants (who were eligible to receive the non-infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size has been chosen to provide very high power (approximately 96%) to establish non-inferiority assuming that the true proportion hospitalized/dead for both the non-infused investigational agent and the active comparator agent is 2.3%, while also providing high power (approximately 85%) assuming that the true proportion hospitalized/dead for the non-infused investigational agent is 0.5% worse, i.e., 2.8%, than the active comparator agent. The rationale for the 2.3% rate for the active comparator agent and for having no adjustment for loss to follow-up is the same as described above in justifying the sample size for infused investigational agents.

The potential power of the study for non-infused agents was evaluated in the same two ways as described above for infused investigational agents using the PASS version 15 sample size calculation software. Use of the binomial enumeration method not taking account of interim analyses gave a power of 96.6% if the non-infused investigational agent and active comparator agent had the same true rates of hospitalization/death (2.3%), and 85.2% power if the non-infused investigational agent had a slightly lower true rate than the active comparator agent (2.8% versus 2.3%). The second (simulation-based) approach did not use a binomial enumeration but taking account of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 96.2% (width of 95% confidence interval around this simulation-based value was 0.08%) if the non-infused investigational agent and the active comparator agent has the same true rate of hospitalization/death (2.3%), and 84.2% power (width 0.15%) if the non-infused investigational agent had a slightly lower true rate than the active comparator agent (2.8% versus 2.3%).

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring of the placebo-controlled phase II, the placebo-controlled phase III and the active-controlled phase III. Additional details on phase II monitoring can be found in protocol version 7.0 section 10.5, and in protocol version 7.0 Appendix

V for placebo-controlled phase III monitoring. Details on active-controlled phase III monitoring are taken for protocol version 7.0 section 10.5.2 as amended in letter of amendment 1 to protocol version 7.0. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to study product or if two participants experience a Grade 4 AE deemed related to study product, enrollment to the study product group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II – Placebo-Controlled Superiority

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III – Placebo-Controlled Superiority

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

2.5.3 Phase III – Active-Controlled Non-Inferiority

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if there is persuasive evidence of non-inferiority (or superiority or inferiority) of an investigational agent versus the active comparator agent in its effect on the hospitalization/death outcome. It is not intended, however, to terminate evaluation of an agent

early for efficacy based on symptom outcome measures. The DSMB may also recommend termination or modification of the study if it appears futile on statistical or operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, three interim analyses for DSMB review are planned for each investigational agent, after approximately 25%, 50% and 75% of the planned enrollment for an investigational agent has been completed and followed through to day 28. At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guidelines for Efficacy or Lack of Efficacy

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses) for the absolute difference (investigational agent minus active comparator agent) in the proportion of participants hospitalized or dead by day 28, relative to thresholds defining non-inferiority, superiority or inferiority of the investigational agent as follows (in the order given):

- The DSMB may recommend releasing results evaluating the effect of an investigational agent when both non-inferiority and superiority of that agent is established based on the confidence interval being entirely below 0% (i.e., supportive of a lower true proportion being hospitalized or dying on the investigational agent than the active comparator agent). If this occurs, consideration will need to be given to the ongoing appropriateness of the active comparator agent as a control for evaluating other investigational agents in the study.

- Early stopping and/or release of results based on non-inferiority should be considered on an agent-by-agent basis. For non-infused agents, the DSMB may recommend releasing results evaluating the effect of an investigational agent when non-inferiority (but not superiority) of that agent is established based on the confidence interval being entirely below 3% (but not entirely below 0%). However, in the interests of also having an adequate safety database for the investigational agent, it is not intended that this recommendation be made before approximately 400 participants have been randomized to receive the agent (or some other number of participants specified in the agent-specific appendix). In addition, the study may continue randomizing participants to the investigational agent in the interests of increasing precision in evaluating the agent; this decision will be made by the study team and sponsor on an agent-by-agent basis. For infused agents, early stopping and/or release of results for non-inferiority should not be considered.

- The DSMB may recommend releasing results and terminating randomization to an investigational agent if inferiority of that agent is established based on the confidence interval being entirely above 0% (i.e., suggesting a higher true proportion being hospitalized or dying on the investigational agent than the active comparator agent).

Examples of how this criterion might be met when evaluating an infused agent and when the observed control rate is close to 2.3% include observing 18/150 versus 3/150 (observed difference 10.0%) at the first interim analysis; 23/300 versus 7/300 (observed difference 5.3%) at the second interim analysis; and 24/450 versus 10/450 at the third interim analysis (observed difference 3.1%). In these examples, all observed differences are higher than the non-inferiority margin of 3%, and are indicative also of the futility of continuing evaluation of the infused investigational agent to demonstrate non-inferiority.

2.6 Graduation to Phase III

The following applies to investigational agents that have been not assessed for graduation to phase III under prior versions of the protocol (version 1.0 to 6.0); note BRII-196+BRII-198 was previously assessed for graduation to phase III under protocol versions 2.0 and 3.0 (clinical sites were enrolling participants under both versions at the time of the graduation analysis).

Each investigational agent that is being considered for evaluation in phase III will be evaluated for safety, for activity in reducing COVID-19 symptoms and hospitalization/death, and for activity in reducing SARS-CoV-2 RNA shedding. An analysis to determine if an agent should graduate from phase II, and enter phase III, will be conducted when 220 participants assigned to the agent or concurrent placebo in phase II evaluation have completed their Day 7 evaluations and have the required data available in the database. Additional interim graduations may be assessed for some agents, see agent-specific appendices in protocol version 7.0 for details.

The DSMB will review unblinded data and make recommendations to NIAID (as trial sponsor) and to the TOC, indicating whether graduation criteria have been met. The recommendation for an agent to enter phase III evaluation will be made by the TOC in discussion with the collaborating company; the collaborating company that is responsible for the agent will decide whether or not to adopt the recommendation. The TOC and collaborating company will also consider which dose to recommend for evaluation in phase III, for investigational agents with more than one dose under evaluation in phase II. NIAID/DAIDS, as the sponsor of the study, will make the final determination regarding graduation of the study product.

The TOC may recommend an agent move directly into phase III, without evaluation in phase II in ACTIV-2, if there is sufficient safety and efficacy data supporting phase III evaluation available from outside of the trial. These agents will not undergo graduation analyses.

Graduation criteria and statistical considerations are discussed in the Graduation Rules SAP.

3 Outcome Measures

All outcome measures are copied from the protocol version 7.0 (including letter of amendment 1). Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the number of days from start of investigational treatment to the first of two consecutive days when any symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, and any symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: At each of days 3, 7 and 14 quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs.
   [For Primary Objective 3]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
[Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2 and for Secondary Objective 8]

   Duration defined as the number of days from start of investigational treatment to the first of four consecutive days when all symptoms are scored as absent. Targeted symptoms as defined in the primary phase II clinical (symptom duration) outcome.

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of two consecutive days that a participant reported return to usual (pre-COVID) health.

7) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

   Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

8) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention, hospitalization, and death. [For Secondary Objective 1].

   Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

9) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or comparator intervention. [For Secondary Objective 2]

10) Phase II only: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

11) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

<u>Virology</u>

12) Phase III (Active-Controlled) only: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs at day 3. [Support of Primary Objective 3]

13) Phase II and III: Level (quantitative) of SARS-CoV-2 RNA from staff-collected NP swabs at days 3, 7, and 14 in phase II and at day 3 in phase III.8.
[For Secondary Objective 3]

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from staff-collected NP swabs at days 0, 3, 7, and 14. [Supportive of both Primary Objective 3 and Secondary Objective 3]

Efficacy

15) Phase II only: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

17) Phase II and III: Death due to any cause or hospitalization due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death due to any cause or hospitalization due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

19) Phase II and III: Death due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

20) Phase II and III: Death due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

21) Phase III: Death due to any cause or hospitalization due to any cause, excluding hospitalizations that are deemed unrelated to COVID-19, during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 9]

    Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or comparator intervention. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
   [For Exploratory Objective 4, with follow-up beyond day 28]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;

Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to its comparator intervention (placebo or active comparator, as appropriate for the agent and phase of evaluation).

- Treated Population: All participants in the Randomized Population who received any investigational agent or its comparator agent (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus its comparator intervention. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to the comparator intervention, who were also eligible to have received the investigational agent of interest. For the placebo-controlled trials, the comparison group will pool across all relevant placebos (i.e. including the placebo for the agent of interest and the placebos for other agents). For the primary placebo-controlled analysis of a specific investigational agent, a supplemental analysis may be undertaken that restricts the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the

evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/comparator intervention occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/comparator intervention.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24).

Week 72: Key visit for evaluation longer-term efficacy and safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLOQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

Virology results generated from specimens with the following conditions reported in the database will be excluded from analyses:

- Thawed
- Invalid Specimen
- Quantity Not Sufficient
- Destroyed

NOTE: Samples with the condition code ‘NOT’ were also to be excluded per the trial sponsor but this code indicates that the specimen was not tested. Thus, no result is expected and no exclusion is needed.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or comparator intervention are not followed.

Participants who have protocol violations, such as those who start investigational agent or comparator intervention outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at “higher” risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the placebo-controlled superiority design. Further details are provided after the table.

<table>
<tr><td colspan="3">Phase III Primary Objective for Efficacy—Placebo-Controlled Superiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.</td></tr>
<tr><td>Estimand description</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td colspan="2">Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td colspan="2">Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td colspan="2">Treated Population</td></tr>
<tr><td>Variable(s)</td><td colspan="2">Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td colspan="2">Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td colspan="2">Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td colspan="2">Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td colspan="2">Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="3">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5).</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (due to any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were

last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17,18, 19, 20 and 21, , evaluate the proportion of participants who are hospitalized or died through week 24, the proportion who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days) or through 504 days (i.e. 72 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher's exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)

3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the active-controlled non-inferiority design. Further details are provided after the table.

Phase III Primary Objective for Efficacy—Active-Controlled Non-Inferiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.	
Estimand description	Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.
Treatment	Investigational agent or active comparator agent (casirivimab and imdevimab).

Target population	Analysis set (analysis population)
Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry	Treated Population
Variable(s)	**Outcome measure(s)**
Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent (coded as 1 if participant died or was hospitalized, and 0 otherwise).	Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent.
Handling of intercurrent events	**Handling of missing data**
None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of the agent to which they were randomized.	Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as not having an event after the date last known to be alive.
Population-level summary measure	**Analysis approach**
Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization over 28 days.	Difference (for investigational agent minus active comparator agent) of the proportion dying or being hospitalized at day 28. See text for further details.

Analysis Approach

The analysis of the primary efficacy outcome in phase III will evaluate the absolute difference in proportion of participants hospitalized (due to any cause) or died (due to any cause), from day 0 through day 28, between randomized arms; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included.

Inference will be based on constructing a two-sided exact 95% confidence interval for the absolute difference in proportions (proportion for the investigational agent minus the proportion for the active comparator agent). If this confidence interval is entirely below the non-inferiority margin of 3%, then a conclusion of non-inferiority of the investigational agent compared with the active

comparator agent will provide reasonable evidence that the investigational agent is effective against COVID-19.

The exact 95% confidence interval will be calculated using the method of Chan and Zhang [Biometrics 1999;55:1201-09] as implemented, for example, in StatXact PROC BINOMIAL for SAS [StatXact 12 PROCs for SAS Users Manual. Cytel Inc., Cambridge, MA; 2019]. This method inverts two one-sided hypothesis tests (with one-sided error rate of 0.025 each) to obtain the confidence interval so providing a confidence interval-based method which preserves the type I error rate in establishing non-inferiority to be 0.025. To preserve confidence interval coverage (and type I error rate for assessing non-inferiority) over multiple interim analyses, the confidence interval will be calculated using a “repeated” confidence interval approach with spending of error rate at each interim analysis using the Land and DeMets approach with an O’Brien and Fleming spending function.

In essence, basing the comparison of treatment groups on the simple proportion of participants who were hospitalized or died assumes that participants who are lost to follow-up before 28 days without prior hospitalization were not hospitalized and did not die by 28 days. The decision to use the simple proportion for analysis rather than use, for example, a Kaplan-Meier estimate of the cumulative proportion of participants hospitalized or dying during the first 28 days of follow-up to account for losses to follow-up was taken for multiple reasons. First, in ACTIV-2 and other COVID-19 trials, most hospitalizations and deaths occur during the first two weeks of follow-up and the study has been designed to have regular contact with participants or their secondary contacts so as to maximize ascertainment of hospitalization and death information. Second, loss to follow-up has been low in the ACTIV-2 study: approximately 3% among higher risk participants. Third, with the very low rates of hospitalization/death expected (e.g., 2.3% for the active comparator agent), confidence interval coverage (and type I error rates) are better preserved at their desired levels through the use of exact statistical methods for analyzing proportions than is achieved using asymptotic statistical methods based on Wald-type analyses using Greenwood’s formula to obtain standard errors for Kaplan-Meier estimates. To assess the potential impact of loss to follow-up (assumed to be non-informative) on the interpretation of results, the following sensitivity analyses will be undertaken, repeating the primary analysis repeated with:

(a) a comparison of the simple proportions using a Wald-based confidence interval; and

(b) a comparison of proportions estimated using Kaplan-Meier methods (with censoring of follow-up at the earlier of day 28 and the time that a participant was last known to be alive) using a Wald-based confidence interval with standard error based on Greenwood’s formula.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 16, 17,18, 19, 20 and 21 evaluate the proportion of participants who die through to day 28, the proportion who are hospitalized or died through week 24, the proportion

who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In the sensitivity analyses based on Kaplan-Meier estimates, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days for outcomes through to 24 weeks) or through 504 days (i.e. 72 times 7 days for outcomes through to 72 weeks).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approach outlined for the primary analysis will be implemented for each subgroup. However, these analyses are likely to involve small numbers of events in most or all subgroups and hence have very limited precision. Because of this, any assessment of treatment by subgroup interaction, if undertaken, will be considered exploratory. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/active comparator Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or comparator intervention) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

In placebo-controlled evaluations, because some agents may be administered using injections or infusions and others will not be, the primary safety analysis may be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed in the same manner as the primary safety outcomes.

Additional longer-term safety outcomes may be assessed, see agent-specific appendices for details.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the number of days from start of investigational agent (day 0, pre-treatment) to the first of two consecutive days when all symptoms scored as moderate

or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and comparator intervention arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic

missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to “absent”, and (b) four consecutive days with resolution of all targeted symptoms to “absent” (i.e., secondary outcome measure 5). For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are “absent” on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis maybe repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to comparator intervention to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) ‘Risk of Severe Disease’ Stratification [this may not be pursued for agents which predominantly enrolled participants who were at ‘lower’ risk for severe COVID progression]
5) Age Group (<60, ≥60)

6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, and 14).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as "0" if above LLoQ, and "1" if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or comparator intervention) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be

calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the number of days from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the

number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;
3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study

day. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

Duration of Symptoms

In support of the symptom duration outcome in phase III, the analysis will be repeated using the same approach described in the primary symptom duration analysis for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to "absent." To address secondary objective 8, and in supportive of the symptom duration outcome in phase III, a similar TTE outcome measure will also be examined defined as time to four consecutive days with resolution of all targeted symptoms to "absent," (i.e. secondary outcome measure 5).

Return to Usual Health

The analysis of return to usual health will be repeated using the same approach described above for a similar TTE outcome measures defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2.2 Secondary Virology

The schedule of evaluations in protocol version 7.0 indicates that only NP swabs will collected in both phase II and phase III, and therefore only analyses of SARS-CoV-2 RNA from NP swabs are outlined below. Some agents may have completed enrollment in phase II prior to implementing protocol version 7.0, and therefore may have additional specimens collected for SARS-CoV-2 RNA testing. If analyses of these additional specimens are pursued, then the approach will be as defined in the relevant previous version of the SAP.

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA at day 3 (this is a secondary outcome for the active-controlled phase 3 only)

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from staff-collected NP swabs at entry and day 3.

The proportion of participants with SARS-CoV-2 RNA < LLoQ day 3 will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. The model will include a main effect for treatment and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained by taking the exponential of the treatment parameter estimate (and associated 95% CI). In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ). Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time for staff-collected NP swabs.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

In phase II only, levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs will be analyzed using participant-specific AUCs. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and

above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: max[0, $\log_{10}$(RNA)-$\log_{10}$(LLoQ)], obtained at the scheduled measurement times between and including day 0 and day 14.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 14 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 14 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will also be provided.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent comparator intervention through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent comparator intervention through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent comparator intervention who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with

frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 72.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

Interim analyses of the placebo-controlled superiority phase III evaluation of an agent was finished at the time of finalization of SAP version 7.0. The following from protocol version 7.0 describes the interim analysis considerations for the active-controlled non-inferiority phase III evaluation of an agent.

The two-sided 95% confidence interval mentioned above [see section 2.5.3 of the SAP] will be adjusted for the multiple interim analyses to preserve the confidence interval coverage to at least 95% (this is also referred to as using “repeated” confidence intervals).

The standard Lan and DeMets approach will be used to achieve this, incorporating an O’Brien and Fleming spending function. For simplicity, the information scale for the spending function will be determined as the proportion of the planned enrollment randomized to the investigational agent being evaluated at the time of the interim analysis. As an example, if in practice, the analyses were after exactly 25%, 50%, 75% and 100% of the planned enrollment, then the nominal confidence intervals used to assess efficacy would have coverage 99.9985% at the first analysis, 99.70% at the second analysis, 98.17% at the third analysis and 95.60% at the fourth analysis (these were obtained from PASS software). However, as the O’Brien and Fleming spending function is very conservative at early interim analyses, making stopping very difficult, for the assessment of inferiority of an investigational agent compared to the active comparator agent, an asymmetric approach will be used to reduce the level of evidence required for early stopping in the event that an investigational agent appears inferior to the active comparator agent. Specifically, if a nominal confidence interval with coverage of greater 99.9% at an early interim analysis is suggested by use of the O’Brien and Fleming spending function, then a nominal confidence interval with coverage of 99.9% will be used instead for assessing inferiority of the investigational agent.

The DSMB will also monitor the proportion hospitalized/dead in the active comparator arm as this key parameter, coupled with the non-inferiority margin, underpins the study design. The study is designed assuming that the underlying true proportion of participants on the active comparator agent is 2.3%. This is the proportion (32/1392) observed for high risk participants in the Regeneron COV-2067 trial for the agent (pooling across doses studied in that trial; FDA communication to DAIDS/NIAID). A 95% confidence interval for this proportion is (1.5%, 3.1%).

An assessment of non-inferiority in this study would be more difficult if the proportion of participants on the active comparator agent in this study is somewhat different from that in the Regeneron COV-2067 (e.g., somewhat outside of the range suggested by the confidence interval).

For example, this might arise if variants of SARS-CoV-2 are present in the study population which the active comparator agent is less effective against. Such an issue would undermine the use of a 3% non-inferiority margin in this study. It may however be addressed by focusing the non-inferiority assessment on the subpopulation in this study without such variants (assuming these have been identified), or in establishing superiority of the investigational agent in the overall study population. This may require a larger sample size to maintain power.

Another potential reason for a somewhat different proportion hospitalized/dead on the active comparator agent in this study versus that in the Regeneron COV-2067 study is that this study is likely to enroll in a number of different countries, whereas the Regeneron COV-2067 enrolled primarily in the United States. Aside from possible differences in circulating variants among countries, differences among countries in clinical practice and/or in the availability of hospital care might lead to differences in hospitalization/death rates. The DSMB will monitor descriptive results by country and provide guidance about countries with notably low or high rates of hospitalization/death.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to

admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. **If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28**:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as

"mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued through to protocol version 6.0. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions. All participants enrolled to evaluate BRII-198+BRII-196 were randomized to active agent or placebo and so placebo is mentioned as a the comparator intervention throughout this appendix.

7.1 Randomization Details

Phase III is stratified by time from symptom onset (≤ 5 days versus > 5 days).

7.2 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.3 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 72).

8 Appendix 3: Statistical Considerations for AZD7442 IV

NOTE: AZD7442 IV is only being evaluated in this study in phase II with a placebo control.

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 72).

9 Appendix 4: Statistical Considerations for AZD7742 IM

NOTE: AZD7442 IM is only being evaluated in this study in phase II with a placebo control.

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 72).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

10.1.2 Exploratory Objectives

1) Phase II and III: To determine whether SNG001 reduces severity of cough or shortness of breath or difficulty breathing through study day 28.

2) Phase II and III: To determine whether SNG001 reduces a COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 among individuals who report moderate to severe shortness of breath or difficulty breathing at day 0.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II and III: Area under the curve of *cough* and *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *cough* and *shortness of breath or difficulty breathing* symptoms in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e.

will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate. In addition, Hodges-Lehmann estimate and associated 95% CI for the location shift between the two arms will be provided.

To address SNG001-specific exploratory objective 2, the COVID-19 severity ranking outcome will compared between arms using the same methods outlined for the secondary analysis of this outcome measure, but restricted to be among those with moderate to severe shortness of breath or difficult breathing at day 0,

11 Appendix 6: Statistical Considerations for Camostat

NOTE:camostat is only being evaluated in this study in phase II with a placebo control.

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a separate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, in phase II, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 72).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Phase II/III Placebo-Controlled and Phase III Active-Controlled

Study Components

Version 8.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 7.0 and

Letter of Amendment #1 and #2 to Protocol v7.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

January 24, 2022

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**
VERSION HISTORY **4**
GLOSSARY OF TERMS **7**
1 INTRODUCTION **8**
1.1 Purpose **8**
1.2 Version History of this SAP **8**
2 STUDY OVERVIEW **9**
2.1 Study Design **9**
2.2 Randomization Process **10**
2.3 Study Objectives **11**
2.3.1 Primary Objectives 11
2.3.2 Secondary Objectives 12
2.3.3 Exploratory Objectives 12
2.4 Overview of Sample Size Considerations **13**
2.4.1 Phase II – Placebo-Controlled Superiority 14
2.4.2 Phase III – Placebo-Controlled Superiority Trial 14
2.4.3 Phase III – Active-Controlled Non-Inferiority Trial 14
2.5 Overview of Formal Interim Monitoring **18**
2.5.1 Phase II – Placebo-Controlled Superiority 18
2.5.2 Phase III – Placebo-Controlled Superiority 19
2.5.3 Phase III – Active-Controlled Non-Inferiority 20
2.6 Graduation to Phase III **21**
3 OUTCOME MEASURES **22**
3.1 Primary Outcome Measures: Phase III **22**
3.2 Primary Outcome Measures: Phase II **23**
3.3 Secondary Outcome Measures **23**
3.4 Other Outcome Measures **26**
4 STATISTICAL PRINCIPLES **28**
4.1 General Considerations **28**
5 ANALYSIS APPROACHES **31**
5.1 Analyses of the Primary Objectives **32**
5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation 32
5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation 36
5.1.2 Primary Safety (Phase II and III) 38
5.1.3 Primary Clinical Symptoms (Phase II) 39

5.1.4 Primary Virologic (Phase II) 43
5.2 Analyses of Secondary Objectives 45
5.2.1 Secondary Clinical Symptoms 46
5.2.2 Secondary Virology 51
5.3 Exploratory Analyses 54
5.3.1 New SARS-CoV-2 among Household Contacts 54
5.3.2 Hospitalization Course 54
5.3.3 Resistance Mutations 55
5.4 Interim Analysis Considerations 55

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 57

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 60
7.1 Randomization Details 60
7.2 Secondary Outcome Measures 60
7.3 Analysis Approaches 60

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 61
8.1 Secondary Outcome Measures 61
8.2 Analysis Approaches 61

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 62
9.1 Secondary Outcome Measures 62
9.2 Analysis Approaches 62

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 63
10.1 Objectives 63
10.1.1 Secondary Objectives 63
10.1.2 Exploratory Objectives 64
10.2 Outcome Measures 64
10.2.1 Secondary Outcome Measures 64
10.2.2 Other Outcome Measures 64
10.3 Analysis Approaches 64
10.3.1 Secondary Analyses 64
10.3.2 Exploratory Analyses 65

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 66
11.1 Objectives 66
11.1.1 Secondary Objectives 66
11.1.2 Exploratory Objectives 66
11.2 Outcome Measures 66
11.2.1 Secondary Outcome Measures 66
11.3 Analysis Approaches 66

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 67

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 68
13.1 Secondary Outcome Measures 68
13.2 Analysis Approaches 68

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at ‘lower’ risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., ‘higher’ risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at ‘higher risk’	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	
6.0	Update SAP to address changed implemented in protocol version 7.0 and Letter of Amendment 1. Specifically: - Updated SAP to reflect changes in study objectives, schedules for some evaluations, and associated outcome measures. - Updated SAP to reflect changes in interim monitoring, mainly related to changes in graduation criteria. - Updated this SAP to focus on the placebo-controlled phase III evaluation. (Note that a separate *revision to the* SAP will be prepared for the active-controlled phase III evaluation introduced in protocol version 7.0). [Note that the italicized text was added in version 7.0 of the SAP].	September 13, 2021
7.0	Updated SAP with the following major changes: - To add details that are specific to the active-controlled phase 3 evaluation of agents based on protocol version 7.0 and Letter of Amendment 1. - To edit some text to provide clarity concerning the analysis approaches which are the same regardless of whether a placebo control or active control is involved. In part, to achieve this, the terminology "comparator intervention" is often used. - To replace the previous section 5.4 concerning interim analysis considerations for the placebo-controlled phase III trial (which have been completed) with interim analysis considerations for the active-controlled phase III trial. - To indicate exclusion from analysis of viral shedding results from samples labelled as 'Thawed', 'Destroyed', 'Quantity Not Sufficient' or 'Invalid Specimen' as approved by the trial sponsor. - To focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.	October 24, 2021

8.0	Updated SAP to address changes implemented in Letter of Amendment 2 to protocol version 7.0. Specifically: - Added oxygen saturation as a secondary outcome in phase 3 - For the SNG001 agent, added phase II secondary objectives. - For the SNG001 agent, revised phase III exploratory objectives. - For the SNG001 agent, added details on subgroup analyses. Removed Hodges-Lehmann analysis from all analysis sections as the validity of this analysis is questionable for the type of data being generated in this study for the affected outcome measures.	January 24, 2022

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III placebo-controlled investigations of ACTIV-2/A5401, as well as the phase III active-controlled investigation introduced in protocol version 7.0. This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 stated that enrollment to all agents (except BRII-196+BRII-198 which was already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents. SAP version 5.0 therefore described planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a placebo. In addition, SAP version 5.0 addressed some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

Protocol version 7.0 introduced an open-label non-inferiority phase III design to compare investigational agents to an active-comparator among persons at higher risk of progression to hospitalization or death. This phase III evaluation is separate from the phase II superiority evaluation of agents compared to placebo among persons at lower risk for progression to hospitalization or death. Changes introduced in SAP version 6.0 focused on changes made under protocol version 7.0 (and letter of amendment #1) that related to the placebo-controlled superiority phase II/III design (note: BRII-196+BRII-198 is the only agent that enrolled in the placebo-controlled phase III design). Changes introduced in SAP version 7.0 address the introduction of the active-controlled non-inferiority phase III trial in protocol version 7.0 (and letter of amendment #1). SAP version 7.0 also introduces the exclusion from statistical analysis of results generated from problematic virologic samples based on a decision made by the DAIDS and study team. In addition, section 5.4 concerning interim analysis considerations was revised to replace considerations for the placebo-controlled phase III trial for which DSMB monitoring has been completed with considerations for the active-controlled phase III trial. Finally, adjustments were made to focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.

SAP version 8.0 implements changes made under letter of amendment #2 to protocol version 7.0, which added oxygen saturation outcome for the active-controlled phase III and new phase III secondary and exploratory objectives for the SNG001 agent. In addition, analyses using the Hodges-Lehmann estimate were removed throughout as the validity of these analyses is questionable for the type of data being generated in this study for the affected outcome measures.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 7.0 and letter of amendments 1 and 2.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient phase II and phase III testing of new agents within the same trial infrastructure.

Version 7.0 of the protocol provides for a blinded phase II evaluation of an investigational agent compared to placebo among participants at lower risk for progression to hospitalization or death, regardless of the mode of administration of the agent; for some agents, enrollment to higher risk participants in phase II was allowed under earlier protocol versions. Agents that graduate to phase III (after initiation of this protocol version) will be evaluated in persons at higher risk for progression to hospitalization or death for non-inferiority to an active comparator, the monoclonal antibody combination of casirivimab plus imdevimab (REGEN-COV, Regeneron), which has been

shown to be effective in this population in preventing hospitalization or death. Protocol version 7.0 also provides for continued follow-up of participants enrolled into a placebo-controlled phase III trial evaluating the combination monoclonal antibody agent, BRII-196 + BRII-198.

When two or more agents are being evaluated in the same phase of the study, the trial design includes sharing of the control group (placebo in phase II and active comparator in phase III) for efficient evaluation of each agent. Note that enrollment to BRII-196+BRII-198 did not coincide with enrollment to other agents in phase III.

Eligible participants will have intensive follow-up through day 28, followed by limited follow up through at least week 72 weeks in phase II and phase III.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0) previous versions of the protocol), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

The phase II trial and the active-controlled phase III trial involve different populations and have separate randomizations. However, the structure of the randomization process is the same for each of the two trials, as described in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of concurrently randomized participants on a given investigational agent and on the comparator group for that agent (i.e. combining participants who were eligible to receive the agent but who were randomized to any of the available placebos in phase II or to the active comparator in phase III).

To achieve having a similar number of participants on the active arm and in the pooled comparator group for a given investigational agent, the randomization occurs in two steps within each trial.

The first randomization is to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents in the trial (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). Trial phase for an agent is accounted for in the participant eligibility (i.e. by the classification of their risk for hospitalization/death as lower or higher). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

Immediately following the first randomization, participants are randomized within an Agent Group in the second randomization to the (active) investigational agent or appropriate comparator (the

matching placebo for agents in phase II, or the active comparator for agents in phase III). For a given participant, the probability of assignment to the active agent or comparator in the second randomization depends on the number of agents currently under investigation that the participant was eligible to receive, as phase II and phase III have distinct populations (phase II is restricted to those at lower risk of progression to hospitalization and death, and phase III is restricted to those at higher risk).

Both the first and second randomizations involve blocked stratified randomization. In phase II, both the first and second randomizations are stratified by time from symptom onset (≤5 days vs >5 days), however, in previous versions of the protocol, in which both 'higher' and 'lower' risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group ('higher' vs 'lower'). In the active-controlled phase III trial introduced in protocol version 7.0, both randomization steps are stratified by country. Under previous versions of the protocol for the placebo-controlled phase III, both randomization steps were only stratified by time from symptom onset (≤ 5 days vs > 5 days). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 7.0 (and letter of amendment #1); corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase 2, which will be addressed outside of this SAP. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed separately.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, and 14 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in NP swabs [Protocol Objective 1.2.3].

4) Phase III: To determine the efficacy of the investigational agent to increase the proportion of participants with NP SARS-CoV-2 RNA below the LLoQ at study day 3 [Protocol Objective 1.2.4]

5) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.5].

6) Phase II: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72 [Protocol Objective 1.2.7].

8) Phase III: To evaluate if the investigational agent reduces the time to sustained symptom resolution through study day 28 [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of hospitalization or death through study day 28, excluding hospitalizations that are determined to be unrelated to COVID-19 [Protocol Objective 1.2.9 (introduced in letter of amendment 1 to protocol version 7.0)].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes and differences between investigational agent and control (placebo in phase II and active comparator in phase III)

across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course in those hospitalized [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 7.0. The sample size for the placebo-controlled superiority phase III design was also the same under protocol versions 2.0 to 6.0 (it was originally defined in Appendix IV of protocol version 2.0 for the agent entered into the study under protocol version 2.0) and is currently detailed in Appendix V of protocol version 7.0 for the BRII-196+BRII-198 agent. Details on the sample size for the non-inferiority active-controlled phase III design are from protocol version 7.0 section 10.4.

2.4.1 Phase II – Placebo-Controlled Superiority

For each investigational agent in phase II, the proposed sample size is 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

2.4.2 Phase III – Placebo-Controlled Superiority Trial

The proposed sample size is 842 participants consisting of 421 participants who receive the active agent and 421 participants who are concurrently randomized to placebo control. This sample size includes the enrollment that occurred during the phase II placebo-controlled evaluation of an agent. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size has been chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This is based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.4.3 Phase III – Active-Controlled Non-Inferiority Trial

The active-controlled Phase III trial is focused on a non-inferiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus an active comparator agent, specifically the monoclonal antibody combination of casirivimab plus imdevimab. The non-inferiority margin for the absolute difference in proportion hospitalized/dead is 3% (investigational agent minus active comparator agent); the rationale for

this choice is described in Section 3.1 of protocol version 7.0. Non-inferiority will be considered established if a two-sided exact 95% confidence interval for the absolute difference is entirely below 3%. Details of the construction of the confidence interval are in section 10.6 pf the protocol and are included further below in this SAP.

The sample size differs between infused investigational agents (600 for the investigational agent and 600 for the concurrently randomized active comparator) and non-infused investigational agents (800 per arm instead of 600 per arm). The rationale for this is that there may be broader clinical utility for non-infused agents such that a slightly higher true hospitalization/death rate may be tolerated in clinical practice.

Sample Size Justification for Infused Investigational Agents

For the evaluation of a specific infused investigational agent, the sample size is 1200 participants including approximately 600 participants randomized to receive the infused investigational agent and approximately 600 participants (who were eligible to receive the infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size has been chosen to provide close to 90% power to establish non-inferiority assuming that the true proportion hospitalized/dead for both the infused investigational agent and the active comparator agent is 2.3%. The rate of 2.3% is based on the observed proportion for casirivimab plus imdevimab combining across doses in the subpopulation of the Regeneron COV-2067 clinical trial who met the criteria for being at high risk of progression to hospitalization/death (FDA communication to DAIDS/NIAID, May 2021). No adjustment for loss to follow-up is made in the sample size as the primary analysis will be based on the observed number of hospitalizations divided by the number of participants who initiated study treatment. In addition, the impact of any loss to follow-up is expected to be minimal as there will be regular contact between research site staff and participants (or their secondary contacts) and previous experience in the study and other trials has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up).

The potential power of the study was evaluated in two ways using the PASS version 15 sample size calculation software. Both used a non-inferiority hypothesis testing approach based on use of the Miettinen and Nurminen score test statistic (which is the basis for calculating the confidence interval used for analysis in this study). The first ignored interim monitoring but used a binomial enumeration method to calculate power and type I error rates. Use of the binomial enumeration method takes account of the discreteness of the binomial distribution (rather than using a normal approximation to the binomial distribution) which may be important in the setting of low hospitalization/death probabilities. Using this approach gave a power of 90.2%. The second approach did not use a binomial enumeration but took account of interim analyses using a standard implementation of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 90.0% (width of 95% confidence interval around this simulation-based value was 0.12%). Based on these two approaches, it is anticipated that the study will have close to 90% power to show non-inferiority for an infused investigational agent assuming that it truly has the same 2.3% hospitalization/death rate as the active comparator agent.

The PASS software was also used to illustrate how the power of the study might change for various scenarios which differ from the scenario assumed (see Table 2.4.3-1). This was undertaken using the first of the two approaches for evaluating prior mentioned above (i.e., using the binomial enumeration approach). Looking at the top part of the table in which both the infused investigational agent and the active comparator agent have the same underlying true hospitalization/ death rate, the power is decreased if the true rate is above the assumed 2.3%, but increased if the true rate is less than 2.3%. If the true rate is 3%, then the power is still above 80%, but if the true rate is 4% it is reduced to 73%.

The middle and lower parts of the table show scenarios in which the infused investigational agent has a true hospitalization/death rate of 0.5% or 1% worse than the active comparator agent, respectively. If the true rate for the active comparator agent is 2.3% and is 2.8% for the infused investigational agent (i.e., 0.5% worse), then the power is reduced to 73%. If the true rate for the active comparator agent is 2.3% and is 3.3% for the infused investigational agent (i.e., 1% worse), then the power is reduced to 50%

Table 2.4.3-1: Power for various scenarios based on non-inferiority hypothesis testing using the likelihood score test statistic (Miettinen and Nurminen method) with binomial enumeration of power and Type I error rate. All scenarios use a 3% non-inferiority margin and one-sided Type-I error rate of 0.025 with a sample size of 600 participants receiving an infused investigational agent and 600 participants receiving the active comparator agent. Power in practice will be slightly reduced from the values shown due to interim monitoring.

Same Underlying True Hospitalization/Death Rate For Active Comparator Agent and Infused Investigational Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
99.4%	1%	1%	2.2%
97.3%	1.5%	1.5%	2.2%
93.2%	2%	2%	2.3%
90.2%	2.3%	2.3%	2.4%
88.1%	2.5%	2.5%	2.4%
83.1%	3%	3%	2.4%
78.1%	3.5%	3.5%	2.4%
73.2%	4%	4%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 0.5% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
92.5%	1%	1.5%	2.2%
85.2%	1.5%	2%	2.2%
77.4%	2%	2.5%	2.3%
73.1%	2.3%	2.8%	2.4%
70.5%	2.5%	3%	2.4%
64.8%	3%	3.5%	2.4%
59.6%	3.5%	4%	2.4%
55.0%	4%	4.5%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 1% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
71.3%	1%	2%	2.2%
61.7%	1.5%	2.5%	2.2%
54.0%	2%	3%	2.3%
50.4%	2.3%	3.3%	2.4%
48.4%	2.5%	3.5%	2.4%
44.0%	3%	4%	2.4%
40.0%	3.5%	4.5%	2.4%
36.7%	4%	5%	2.4%
*Actual type I error rate is slightly lower than assumed rate of 2.5% because of discreteness of the binomial distribution.			

Sample Size Justification for Non-infused Investigational Agents

For the evaluation of a specific non-infused investigational agent, the sample size will include approximately 800 participants randomized to receive the non-infused investigational agent and approximately 800 participants (who were eligible to receive the non-infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size has been chosen to provide very high power (approximately 96%) to establish non-inferiority assuming that the true proportion hospitalized/dead for both the non-infused investigational agent

and the active comparator agent is 2.3%, while also providing high power (approximately 85%) assuming that the true proportion hospitalized/dead for the non-infused investigational agent is 0.5% worse, i.e., 2.8%, than the active comparator agent. The rationale for the 2.3% rate for the active comparator agent and for having no adjustment for loss to follow-up is the same as described above in justifying the sample size for infused investigational agents.

The potential power of the study for non-infused agents was evaluated in the same two ways as described above for infused investigational agents using the PASS version 15 sample size calculation software. Use of the binomial enumeration method not taking account of interim analyses gave a power of 96.6% if the non-infused investigational agent and active comparator agent had the same true rates of hospitalization/death (2.3%), and 85.2% power if the non-infused investigational agent had a slightly lower true rate than the active comparator agent (2.8% versus 2.3%). The second (simulation-based) approach did not use a binomial enumeration but taking account of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 96.2% (width of 95% confidence interval around this simulation-based value was 0.08%) if the non-infused investigational agent and the active comparator agent has the same true rate of hospitalization/death (2.3%), and 84.2% power (width 0.15%) if the non-infused investigational agent had a slightly lower true rate than the active comparator agent (2.8% versus 2.3%).

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring of the placebo-controlled phase II, the placebo-controlled phase III and the active-controlled phase III. Additional details on phase II monitoring can be found in protocol version 7.0 section 10.5, and in protocol version 7.0 Appendix V for placebo-controlled phase III monitoring. Details on active-controlled phase III monitoring are taken for protocol version 7.0 section 10.5.2 as amended in letter of amendment 1 to protocol version 7.0. Statistical considerations for interim monitoring are shown in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to study product or if two participants experience a Grade 4 AE deemed related to study product, enrollment to the study product group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II – Placebo-Controlled Superiority

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III – Placebo-Controlled Superiority

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the

death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

2.5.3 Phase III – Active-Controlled Non-Inferiority

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if there is persuasive evidence of non-inferiority (or superiority or inferiority) of an investigational agent versus the active comparator agent in its effect on the hospitalization/death outcome. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend termination or modification of the study if it appears futile on statistical or operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, three interim analyses for DSMB review are planned for each investigational agent, after approximately 25%, 50% and 75% of the planned enrollment for an investigational agent has been completed and followed through to day 28. At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guidelines for Efficacy or Lack of Efficacy

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses) for the absolute difference (investigational agent minus active comparator agent) in the proportion of participants hospitalized or dead by day 28, relative to thresholds defining non-inferiority, superiority or inferiority of the investigational agent as follows (in the order given):

- The DSMB may recommend releasing results evaluating the effect of an investigational agent when both non-inferiority and superiority of that agent is established based on the confidence interval being entirely below 0% (i.e., supportive of a lower true proportion being hospitalized or dying on the investigational agent than the active comparator agent). If this occurs, consideration will need to be given to the ongoing appropriateness of the active comparator agent as a control for evaluating other investigational agents in the study.

- Early stopping and/or release of results based on non-inferiority should be considered on an agent-by-agent basis. For non-infused agents, the DSMB may recommend releasing results evaluating the effect of an investigational agent when non-inferiority (but not superiority) of that agent is established based on the confidence interval being entirely below 3% (but not entirely below 0%). However, in the interests of also having an adequate safety database for the investigational agent, it is not intended that this recommendation be made before approximately 400 participants have been randomized to receive the agent (or some other number of participants specified in the agent-specific appendix). In addition, the study may continue randomizing participants to the investigational agent in the interests of increasing precision in evaluating the agent; this decision will be made by the study team and sponsor on an agent-by-agent basis. For infused agents, early stopping and/or release of results for non-inferiority should not be considered.

- The DSMB may recommend releasing results and terminating randomization to an investigational agent if inferiority of that agent is established based on the confidence interval being entirely above 0% (i.e., suggesting a higher true proportion being hospitalized or dying on the investigational agent than the active comparator agent). Examples of how this criterion might be met when evaluating an infused agent and when the observed control rate is close to 2.3% include observing 18/150 versus 3/150 (observed difference 10.0%) at the first interim analysis; 23/300 versus 7/300 (observed difference 5.3%) at the second interim analysis; and 24/450 versus 10/450 at the third interim analysis (observed difference 3.1%). In these examples, all observed differences are higher than the non-inferiority margin of 3%, and are indicative also of the futility of continuing evaluation of the infused investigational agent to demonstrate non-inferiority.

2.6 Graduation to Phase III

The following applies to investigational agents that have been not assessed for graduation to phase III under prior versions of the protocol (version 1.0 to 6.0); note BRII-196+BRII-198 was

previously assessed for graduation to phase III under protocol versions 2.0 and 3.0 (clinical sites were enrolling participants under both versions at the time of the graduation analysis).

Each investigational agent that is being considered for evaluation in phase III will be evaluated for safety, for activity in reducing COVID-19 symptoms and hospitalization/death, and for activity in reducing SARS-CoV-2 RNA shedding. An analysis to determine if an agent should graduate from phase II, and enter phase III, will be conducted when 220 participants assigned to the agent or concurrent placebo in phase II evaluation have completed their Day 7 evaluations and have the required data available in the database. Additional interim graduations may be assessed for some agents, see agent-specific appendices in protocol version 7.0 for details.

The DSMB will review unblinded data and make recommendations to NIAID (as trial sponsor) and to the TOC, indicating whether graduation criteria have been met. The recommendation for an agent to enter phase III evaluation will be made by the TOC in discussion with the collaborating company; the collaborating company that is responsible for the agent will decide whether or not to adopt the recommendation. The TOC and collaborating company will also consider which dose to recommend for evaluation in phase III, for investigational agents with more than one dose under evaluation in phase II. NIAID/DAIDS, as the sponsor of the study, will make the final determination regarding graduation of the study product.

The TOC may recommend an agent move directly into phase III, without evaluation in phase II in ACTIV-2, if there is sufficient safety and efficacy data supporting phase III evaluation available from outside of the trial. These agents will not undergo graduation analyses.

Graduation criteria and statistical considerations are discussed in the Graduation Rules SAP.

3 Outcome Measures

All outcome measures are copied from the protocol version 7.0 (including letter of amendment 1). Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the number of days from start of investigational treatment to the first of two consecutive days when any symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, and any symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: At each of days 3, 7 and 14 quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs.
   [For Primary Objective 3]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
   [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2]

Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2 and for Secondary Objective 8]

Duration defined as the number of days from start of investigational treatment to the first of four consecutive days when all symptoms are scored as absent. Targeted symptoms as defined in the primary phase II clinical (symptom duration) outcome.

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of two consecutive days that a participant reported return to usual (pre-COVID) health.

7) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

8) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention, hospitalization, and death. [For Secondary Objective 1].

Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first

day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

9) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or comparator intervention. [For Secondary Objective 2]

10) Phase II and Phase III: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

11) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

12) Phase III (Active-Controlled) only: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs at day 3. [Support of Primary Objective 3]

13) Phase II and III: Level (quantitative) of SARS-CoV-2 RNA from staff-collected NP swabs at days 3, 7, and 14 in phase II and at day 3 in phase III.8.
[For Secondary Objective 3]

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from staff-collected NP swabs at days 0, 3, 7, and 14. [Supportive of both Primary Objective 3 and Secondary Objective 3]

Efficacy

15) Phase II only: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

17) Phase II and III: Death due to any cause or hospitalization due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death due to any cause or hospitalization due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

19) Phase II and III: Death due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

20) Phase II and III: Death due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

21) Phase III: Death due to any cause or hospitalization due to any cause, excluding hospitalizations that are deemed unrelated to COVID-19, during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 9]

    Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic. An Adjudication committee is evaluating the relatedness of hospitalization due to COVID-19.

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or comparator intervention. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to its comparator intervention (placebo or active comparator, as appropriate for the agent and phase of evaluation).

- Treated Population: All participants in the Randomized Population who received any investigational agent or its comparator agent (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus its comparator intervention. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to the comparator intervention, who were also eligible to have received the investigational agent of interest. For the placebo-controlled trials, the comparison group will pool across all relevant placebos (i.e. including the placebo for the agent of interest and the placebos for other agents). For the primary placebo-controlled analysis of a specific investigational agent, a supplemental analysis may be undertaken that restricts the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/comparator intervention occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/comparator intervention.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24).

Week 72: Key visit for evaluation longer-term efficacy and safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLoQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

Virology results generated from specimens with the following conditions reported in the database will be excluded from analyses:

- Thawed;
- Invalid Specimen;
- Quantity Not Sufficient;
- Destroyed.

Note: Samples with the condition code ‘NOT’ were also to be excluded per the trial sponsor but this code indicates that the specimen was not tested. Thus, no result is expected and no exclusion is needed.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or comparator intervention are not followed.

Participants who have protocol violations, such as those who start investigational agent or comparator intervention outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at “higher” risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the placebo-controlled superiority design. Further details are provided after the table.

<table>
<tr><td colspan="2">Phase III Primary Objective for Efficacy—Placebo-Controlled Superiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.</td></tr>
<tr><td>Estimand description</td><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td>Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td>Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td>Treated Population</td></tr>
<tr><td>Variable(s)</td><td>Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td>Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td>Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td>Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td>Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td>Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="2">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5).</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (due to any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17, 18, 19, 20 and 21, evaluate the proportion of participants who are hospitalized or died through week 24, the proportion who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days) or through 504 days (i.e. 72 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher’s exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis

will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Site (if applicable) or site location (if applicable)

   Subgroup analyses by site will be considered if there are a limited number of sites that contributed to enrollment. Otherwise, subgroup analyses by site location (e.g. by country or region) will be conducted if non-US sites contribute to enrollment.

5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the active-controlled non-inferiority design. Further details are provided after the table.

Phase III Primary Objective for Efficacy—Active-Controlled Non-Inferiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.	
Estimand description	Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.
Treatment	Investigational agent or active comparator agent (casirivimab and imdevimab).

Target population	Analysis set (analysis population)
Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry	Treated Population
Variable(s)	**Outcome measure(s)**
Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent (coded as 1 if participant died or was hospitalized, and 0 otherwise).	Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent.
Handling of intercurrent events	**Handling of missing data**
None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of the agent to which they were randomized.	Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as not having an event after the date last known to be alive.
Population-level summary measure	**Analysis approach**
Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization over 28 days.	Difference (for investigational agent minus active comparator agent) of the proportion dying or being hospitalized at day 28. See text for further details.

Analysis Approach

The analysis of the primary efficacy outcome in phase III will evaluate the absolute difference in proportion of participants hospitalized (due to any cause) or died (due to any cause), from day 0 through day 28, between randomized arms; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included.

Inference will be based on constructing a two-sided exact 95% confidence interval for the absolute difference in proportions (proportion for the investigational agent minus the proportion for the active comparator agent). If this confidence interval is entirely below the non-inferiority margin of 3%, then a conclusion of non-inferiority of the investigational agent compared with the active

comparator agent will provide reasonable evidence that the investigational agent is effective against COVID-19.

The exact 95% confidence interval will be calculated using the method of Chan and Zhang [Biometrics 1999;55:1201-09] as implemented, for example, in StatXact PROC BINOMIAL for SAS [StatXact 12 PROCs for SAS Users Manual. Cytel Inc., Cambridge, MA; 2019]. This method inverts two one-sided hypothesis tests (with one-sided error rate of 0.025 each) to obtain the confidence interval so providing a confidence interval-based method which preserves the type I error rate in establishing non-inferiority to be 0.025. To preserve confidence interval coverage (and type I error rate for assessing non-inferiority) over multiple interim analyses, the confidence interval will be calculated using a “repeated” confidence interval approach with spending of error rate at each interim analysis using the Land and DeMets approach with an O’Brien and Fleming spending function.

In essence, basing the comparison of treatment groups on the simple proportion of participants who were hospitalized or died assumes that participants who are lost to follow-up before 28 days without prior hospitalization were not hospitalized and did not die by 28 days. The decision to use the simple proportion for analysis rather than use, for example, a Kaplan-Meier estimate of the cumulative proportion of participants hospitalized or dying during the first 28 days of follow-up to account for losses to follow-up was taken for multiple reasons. First, in ACTIV-2 and other COVID-19 trials, most hospitalizations and deaths occur during the first two weeks of follow-up and the study has been designed to have regular contact with participants or their secondary contacts so as to maximize ascertainment of hospitalization and death information. Second, loss to follow-up has been low in the ACTIV-2 study: approximately 3% among higher risk participants. Third, with the very low rates of hospitalization/death expected (e.g., 2.3% for the active comparator agent), confidence interval coverage (and type I error rates) are better preserved at their desired levels through the use of exact statistical methods for analyzing proportions than is achieved using asymptotic statistical methods based on Wald-type analyses using Greenwood’s formula to obtain standard errors for Kaplan-Meier estimates. To assess the potential impact of loss to follow-up (assumed to be non-informative) on the interpretation of results, the following sensitivity analyses will be undertaken, repeating the primary analysis repeated with:

(a) a comparison of the simple proportions using a Wald-based confidence interval; and

(b) a comparison of proportions estimated using Kaplan-Meier methods (with censoring of follow-up at the earlier of day 28 and the time that a participant was last known to be alive) using a Wald-based confidence interval with standard error based on Greenwood’s formula.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 16, 17, 18, 19, 20 and 21 evaluate the proportion of participants who die through to day 28, the proportion who are hospitalized or died through week 24, the proportion

who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In the sensitivity analyses based on Kaplan-Meier estimates, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days for outcomes through to 24 weeks) or through 504 days (i.e. 72 times 7 days for outcomes through to 72 weeks).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approach outlined for the primary analysis will be implemented for each subgroup. However, these analyses are likely to involve small numbers of events in most or all subgroups and hence have very limited precision. Because of this, any assessment of treatment by subgroup interaction, if undertaken, will be considered exploratory. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/active comparator Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or comparator intervention) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

In placebo-controlled evaluations, because some agents may be administered using injections or infusions and others will not be, the primary safety analysis may be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed in the same manner as the primary safety outcomes.

Additional longer-term safety outcomes may be assessed, see agent-specific appendices for details.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the number of days from start of investigational agent (day 0, pre-treatment) to the first of two consecutive days when all symptoms scored as moderate

or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and comparator intervention arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic

missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to “absent”, and (b) four consecutive days with resolution of all targeted symptoms to “absent” (i.e., secondary outcome measure 5). For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are “absent” on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis maybe repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to comparator intervention to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) ‘Risk of Severe Disease’ Stratification [this may not be pursued for agents which predominantly enrolled participants who were at ‘lower’ risk for severe COVID progression]
5) Age Group (<60, ≥60)

6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, and 14).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as "0" if above LLoQ, and "1" if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or comparator intervention) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the number of days from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;
3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model

may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

Duration of Symptoms

In support of the symptom duration outcome in phase III, the analysis will be repeated using the same approach described in the primary symptom duration analysis for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to “absent.” To address secondary objective 8, and in supportive of the symptom duration outcome in phase III, a similar TTE outcome measure will also be examined defined as time to four consecutive days with resolution of all targeted symptoms to “absent,” (i.e. secondary outcome measure 5).

Return to Usual Health

The analysis of return to usual health will be repeated using the same approach described above for a similar TTE outcome measures defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2.2 Secondary Virology

The schedule of evaluations in protocol version 7.0 indicates that only NP swabs will collected in both phase II and phase III, and therefore only analyses of SARS-CoV-2 RNA from NP swabs are outlined below. Some agents may have completed enrollment in phase II prior to implementing protocol version 7.0, and therefore may have additional specimens collected for SARS-CoV-2 RNA testing. If analyses of these additional specimens are pursued, then the approach will be as defined in the relevant previous version of the SAP.

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA at day 3 (this is a secondary outcome for the active-controlled phase 3 only)

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from staff-collected NP swabs at entry and day 3.

The proportion of participants with SARS-CoV-2 RNA < LLoQ day 3 will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. The model will include a main effect for treatment and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained by taking the exponential of the treatment parameter estimate (and associated 95% CI). In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ). Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time for staff-collected NP swabs.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

In phase II only, levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs will be analyzed using participant-specific AUCs. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and

above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: max[0, $\log_{10}$(RNA)-$\log_{10}$(LLoQ)], obtained at the scheduled measurement times between and including day 0 and day 14.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 14 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 14 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent comparator intervention through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent comparator intervention through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent comparator intervention who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with

frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 72.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

Interim analyses of the placebo-controlled superiority phase III evaluation of an agent was finished at the time of finalization of SAP version 7.0. The following from protocol version 7.0 describes the interim analysis considerations for the active-controlled non-inferiority phase III evaluation of an agent.

The two-sided 95% confidence interval mentioned above [see section 2.5.3 of the SAP] will be adjusted for the multiple interim analyses to preserve the confidence interval coverage to at least 95% (this is also referred to as using “repeated” confidence intervals).

The standard Lan and DeMets approach will be used to achieve this, incorporating an O’Brien and Fleming spending function. For simplicity, the information scale for the spending function will be determined as the proportion of the planned enrollment randomized to the investigational agent being evaluated at the time of the interim analysis. As an example, if in practice, the analyses were after exactly 25%, 50%, 75% and 100% of the planned enrollment, then the nominal confidence intervals used to assess efficacy would have coverage 99.9985% at the first analysis, 99.70% at the second analysis, 98.17% at the third analysis and 95.60% at the fourth analysis (these were obtained from PASS software). However, as the O’Brien and Fleming spending function is very conservative at early interim analyses, making stopping very difficult, for the assessment of inferiority of an investigational agent compared to the active comparator agent, an asymmetric approach will be used to reduce the level of evidence required for early stopping in the event that an investigational agent appears inferior to the active comparator agent. Specifically, if a nominal confidence interval with coverage of greater 99.9% at an early interim analysis is suggested by use of the O’Brien and Fleming spending function, then a nominal confidence interval with coverage of 99.9% will be used instead for assessing inferiority of the investigational agent.

The DSMB will also monitor the proportion hospitalized/dead in the active comparator arm as this key parameter, coupled with the non-inferiority margin, underpins the study design. The study is designed assuming that the underlying true proportion of participants on the active comparator agent is 2.3%. This is the proportion (32/1392) observed for high risk participants in the Regeneron COV-2067 trial for the agent (pooling across doses studied in that trial; FDA communication to DAIDS/NIAID). A 95% confidence interval for this proportion is (1.5%, 3.1%).

An assessment of non-inferiority in this study would be more difficult if the proportion of participants on the active comparator agent in this study is somewhat different from that in the Regeneron COV-2067 (e.g., somewhat outside of the range suggested by the confidence interval).

For example, this might arise if variants of SARS-CoV-2 are present in the study population which the active comparator agent is less effective against. Such an issue would undermine the use of a 3% non-inferiority margin in this study. It may however be addressed by focusing the non-inferiority assessment on the subpopulation in this study without such variants (assuming these have been identified), or in establishing superiority of the investigational agent in the overall study population. This may require a larger sample size to maintain power.

Another potential reason for a somewhat different proportion hospitalized/dead on the active comparator agent in this study versus that in the Regeneron COV-2067 study is that this study is likely to enroll in a number of different countries, whereas the Regeneron COV-2067 enrolled primarily in the United States. Aside from possible differences in circulating variants among countries, differences among countries in clinical practice and/or in the availability of hospital care might lead to differences in hospitalization/death rates. The DSMB will monitor descriptive results by country and provide guidance about countries with notably low or high rates of hospitalization/death.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. **If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28**:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during

hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued through to protocol version 6.0. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions. All participants enrolled to evaluate BRII-198+BRII-196 were randomized to active agent or placebo and so placebo is mentioned as the comparator intervention throughout this appendix.

7.1 Randomization Details

Phase III is stratified by time from symptom onset (≤ 5 days versus > 5 days).

7.2 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.3 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 72).

8 Appendix 3: Statistical Considerations for AZD7442 IV

NOTE: AZD7442 IV is only being evaluated in this study in phase II with a placebo control.

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 72).

9 Appendix 4: Statistical Considerations for AZD7742 IM

NOTE: AZD7442 IM is only being evaluated in this study in phase II with a placebo control.

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 72).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

2) Phase III: To determine whether SNG001 reduces hospitalization or death through study day 28 among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

3) Phase III: To determine whether SNG001 reduces duration of targeted COVID-19-associated symptoms through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

4) Phase III: To determine whether SNG001 reduces COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

5) Phase III: To determine whether SNG001 reduces progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at entry, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

6) Phase III: To determine whether SNG001 increases proportion of individuals with pulse oximetry measurement of ≥ 96% through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

7) Phase III: To determine whether SNG001 reduces the time to sustained symptom resolution through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

8) Phase III: To determine whether SNG001 prevents the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72, among individuals in the subgroup who report moderate or severe shortness of breath or

difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

9) Phase III: To determine whether SNG001 prevents the composite endpoint of hospitalization or death through stay day 28, excluding hospitalizations that are determined to be unrelated to COVID-19, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

10.1.2 Exploratory Objectives

1) Phase II and III: To determine whether SNG001 reduces severity of cough or shortness of breath or difficulty breathing through study day 28.

2) Phase II: To determine whether SNG001 reduces a COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 among individuals who report moderate to severe shortness of breath or difficulty breathing at day 0.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II: Area under the curve of *cough* and *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *cough* and *shortness of breath or difficulty breathing* symptoms in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

10.3.1 Secondary Analyses

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in

a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

The secondary objectives addressing analyses among people who reported moderate or severe shortness of breath or difficulty breathing at day 0, and among people who reported severe shortness of breath or difficulty breathing at day 0, will be undertaken in the same manner as the analyses of this outcomes among the overall study population.

10.3.2 Exploratory Analyses

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate.

To address SNG001-specific exploratory objective 2, the COVID-19 severity ranking outcome will compared between arms using the same methods outlined for the secondary analysis of this outcome measure, but restricted to be among those with moderate to severe shortness of breath or difficult breathing at day 0,

11 Appendix 6: Statistical Considerations for Camostat

Note: camostat is only being evaluated in this study in phase II with a placebo control.

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a separate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, in phase II, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 72).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Phase II/III Placebo-Controlled and Phase III Active-Controlled

Study Components

Version 9.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 8.0

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

February 25, 2022

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **5**

GLOSSARY OF TERMS **8**

1 INTRODUCTION **9**
1.1 Purpose **9**
1.2 Version History of this SAP **9**

2 STUDY OVERVIEW **11**
2.1 Study Design **11**
2.2 Randomization Process **12**
2.3 Study Objectives **13**
2.3.1 Primary Objectives 13
2.3.2 Secondary Objectives 14
2.3.3 Exploratory Objectives 14
2.4 Overview of Sample Size Considerations **15**
2.4.1 Phase II – Placebo-Controlled Superiority 16
2.4.2 Phase III – Placebo-Controlled Superiority Trial Used for the Evaluation of BRII-196+BRII-198 16
2.4.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant) 17
2.4.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 19
2.5 Overview of Formal Interim Monitoring **22**
2.5.1 Phase II – Placebo-Controlled Superiority 22
2.5.2 Phase III – Placebo-Controlled Superiority Used for the Evaluation of BRII-196+BRII-198 23
2.5.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant) 24
2.5.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 26
2.6 Graduation to Phase III **27**

3 OUTCOME MEASURES **28**
3.1 Primary Outcome Measures: Phase III **28**
3.2 Primary Outcome Measures: Phase II **28**
3.3 Secondary Outcome Measures **29**

3.4 Other Outcome Measures 32

4 STATISTICAL PRINCIPLES 34
4.1 General Considerations 34

5 ANALYSIS APPROACHES 37
5.1 Analyses of the Primary Objectives 38
5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation 38
5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation 42
5.1.2 Primary Safety (Phase II and III) 44
5.1.3 Primary Clinical Symptoms (Phase II) 45
5.1.4 Primary Virologic (Phase II) 49
5.2 Analyses of Secondary Objectives 51
5.2.1 Secondary Clinical Symptoms 52
5.2.2 Secondary Virology 57
5.3 Exploratory Analyses 60
5.3.1 New SARS-CoV-2 among Household Contacts 60
5.3.2 Hospitalization Course 60
5.3.3 Resistance Mutations 61
5.4 Interim Analysis Considerations 61

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 63

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 66
7.1 Randomization Details 66
7.2 Secondary Outcome Measures 66
7.3 Analysis Approaches 66

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 67
8.1 Secondary Outcome Measures 67
8.2 Analysis Approaches 67

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 68
9.1 Secondary Outcome Measures 68
9.2 Analysis Approaches 68

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 69
10.1 Objectives 69
10.1.1 Secondary Objectives 69
10.1.2 Exploratory Objectives 70
10.2 Outcome Measures 70
10.2.1 Secondary Outcome Measures 70
10.2.2 Other Outcome Measures 70
10.3 Analysis Approaches 70
10.3.1 Secondary Analyses 70
10.3.2 Exploratory Analyses 71

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 72
11.1 Objectives 72
11.1.1 Secondary Objectives 72
11.1.2 Exploratory Objectives 72
11.2 Outcome Measures 72
11.2.1 Secondary Outcome Measures 72
11.3 Analysis Approaches 72

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 73
12.1 Study Design and Analysis Population Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185 73
12.2 Additional Analysis Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185 74
12.3 Analysis Considerations for the Non-Omicron Subpopulation Enrolled under Protocol Version 7.0 75
12.4 Data and Safety Monitoring for the Evaluation of SAB-185 under Protocol Version 8.0 75

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 77
13.1 Secondary Outcome Measures 77
13.2 Analysis Approaches 77

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at ‘lower’ risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., ‘higher’ risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at ‘higher risk’	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	
6.0	Update SAP to address changed implemented in protocol version 7.0 and Letter of Amendment 1. Specifically: - Updated SAP to reflect changes in study objectives, schedules for some evaluations, and associated outcome measures. - Updated SAP to reflect changes in interim monitoring, mainly related to changes in graduation criteria. - Updated this SAP to focus on the placebo-controlled phase III evaluation. (Note that a separate *revision to the* SAP will be prepared for the active-controlled phase III evaluation introduced in protocol version 7.0). [Note that the italicized text was added in version 7.0 of the SAP].	September 13, 2021
7.0	Updated SAP with the following major changes: - To add details that are specific to the active-controlled phase 3 evaluation of agents based on protocol version 7.0 and Letter of Amendment 1. - To edit some text to provide clarity concerning the analysis approaches which are the same regardless of whether a placebo control or active control is involved. In part, to achieve this, the terminology "comparator intervention" is often used. - To replace the previous section 5.4 concerning interim analysis considerations for the placebo-controlled phase III trial (which have been completed) with interim analysis considerations for the active-controlled phase III trial. - To indicate exclusion from analysis of viral shedding results from samples labelled as 'Thawed', 'Destroyed', 'Quantity Not Sufficient' or 'Invalid Specimen' as approved by the trial sponsor. - To focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.	October 24, 2021

8.0	Updated SAP to address changes implemented in Letter of Amendment 2 to protocol version 7.0. Specifically: - Added oxygen saturation as a secondary outcome in phase 3 - For the SNG001 agent, added phase II secondary objectives. - For the SNG001 agent, revised phase III exploratory objectives. - For the SNG001 agent, added details on subgroup analyses. Removed Hodges-Lehmann analysis from all analysis sections as the validity of this analysis is questionable for the type of data being generated in this study for the affected outcome measures.	January 24, 2022
9.0	Updated SAP to address changes implemented in protocol version 8.0. Specifically: - Added design details of new superiority placebo-controlled phase III trial that was introduced, including differences from the phase III superiority evaluation of the BRII agent - Outlined statistical analyses of the SAB-185 agent including implications of changing the phase III design due to the Omicron variant - Changed the Primary Symptom Duration Outcome in phase II as per change in protocol version 8.0 - Corrected the phase II “other” outcome measure for the SNG001 agent	February 25, 2022

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III placebo-controlled investigations of ACTIV-2/A5401, as well as the phase III active-controlled investigation introduced in protocol version 7.0 and the phase III placebo-controlled investigation introduced in protocol version 8.0 . This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 stated that enrollment to all agents (except BRII-196+BRII-198 which was already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents (note that this was subsequently changed in protocol version 8.0—see below). SAP version 5.0 therefore described planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a placebo. In

addition, SAP version 5.0 addressed some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

Protocol version 7.0 introduced an open-label non-inferiority phase III design to compare investigational agents to an active-comparator among persons at higher risk of progression to hospitalization or death. This phase III evaluation was separate from the phase II superiority evaluation of agents compared to placebo among persons at lower risk for progression to hospitalization or death. Changes introduced in SAP version 6.0 focused on changes made under protocol version 7.0 (and letter of amendment #1) that related to the placebo-controlled superiority phase II/III design (note: BRII-196+BRII-198 is the only agent that enrolled in the placebo-controlled phase III design until protocol version 8.0 was introduced when a placebo-controlled design was also used for the phase III evaluation of SAB-185). Changes introduced in SAP version 7.0 addressed the introduction of the active-controlled non-inferiority phase III trial in protocol version 7.0 (and letter of amendment #1). SAP version 7.0 also introduced the exclusion from statistical analysis of results generated from problematic virologic samples based on a decision made by the DAIDS and study team. In addition, section 5.4 concerning interim analysis considerations was revised to replace considerations for the placebo-controlled phase III trial for which DSMB monitoring had been completed with considerations for the active-controlled phase III trial. Finally, adjustments were made to focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.

SAP version 8.0 implemented changes made under letter of amendment #2 to protocol version 7.0, which added oxygen saturation outcome for the active-controlled phase III and new phase III secondary and exploratory objectives for the SNG001 agent. In addition, analyses using the Hodges-Lehmann estimate were removed throughout as the validity of these analyses is questionable for the type of data being generated in this study for the affected outcome measures.

While the phase III evaluation of SAB-185 was ongoing using an active-controlled non-inferiority design, in vitro data suggested that the active control agent would not have activity against the newly emergent Omicron variant of SARS-CoV-2. As a result, enrollment to the phase III non-inferiority trial was terminated and was redesigned in order to continue a phase III evaluation of SAB-185. This design was defined in protocol version 8.0 and provided for phase III evaluation of an investigational agent versus placebo, but allowing participants to receive other COVID-19 treatments after study entry if available (availability was, however, expected to be very limited). In essence, this led to the reintroduction of a placebo-controlled phase III trial and the general approach for statistical analyses in protocol version 8.0 for this phase III trial follows the earlier plan for the placebo-controlled phase III evaluation of BRII-196+BRII-198. SAP version 9.0 was implemented to describe these changes.

Participants infected with the Omicron SARS-CoV-2 variant who were randomized under protocol version 7.0 to the “active” control agent in the phase III non-inferiority evaluation of SAB-185 were thought to have been treated with an ineffective agent, so functionally with a placebo from an efficacy perspective, The SAB-185-specific appendix of protocol version 8.0 therefore specifies that the subpopulation of participants enrolled in the non-inferiority phase III evaluation of SAB-

185 under protocol version 7.0 who were definitely or very likely infected with the Omicron variant would be included in the analysis population for the placebo-controlled evaluation of SAB-185. This particular nuance is described in more detail in the SAB-185-specific appendix of this SAP.

SAP version 9.0 also includes a change to an exploratory objective and associated outcome measure for the evaluation of SNG001 that was introduced in protocol version 7.0 but was not reflected in the applicable previous versions of the SAP.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 8.0 for the placebo-controlled phase II and phase III evaluations of investigational agents. This section also includes a description of the non-inferiority phase III evaluation of investigational agents per protocol version 7.0 and letter of amendments 1 and 2.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient phase II and phase III testing of new agents within the same trial infrastructure.

Version 8.0 of the protocol provides for a blinded phase II evaluation of an investigational agent compared to placebo among participants at lower risk for progression to hospitalization or death, regardless of the mode of administration of the agent; for some agents, enrollment to higher risk participants in phase II was allowed under earlier protocol versions.

Based on protocol-specified criteria, agents could graduate from phase II to phase III evaluation. The phase III evaluation of investigational agents under all protocol versions has been in a population of participants at higher risk of hospitalization or death (though the definition of “higher risk” as changed across protocol versions). When protocol version 8.0 was introduced, only two agents had graduated to phase III evaluation and started enrollment: BRII-196+BRII-198 and SAB-185. For these two agents, protocol version 8.0 provides for:

- Continued follow-up of participants enrolled under protocol versions 2.0 to 6.0 into a placebo-controlled phase III trial evaluating the combination monoclonal antibody agent, BRII-196 + BRII-198.
- Continued follow-up of participants enrolled under protocol version 7.0 into a non-inferiority phase III trial evaluating the polyclonal antibody agent, SAB-185 using the monoclonal antibody combination of casirivimab plus imdevimab (REGEN-COV, Regeneron) as the control regimen. As noted above, enrollment to this non-inferiority trial was terminated because of an anticipated lack of efficacy of casirivimab plus imdevimab against the Omicron SARS-CoV-2 variant that became widely prevalent.

- Enrollment of participants into a placebo-controlled phase III trial evaluating SAB-185. In the latter trial, use of COVID-19 treatments obtained outside of the trial is allowed in both randomized arms, if available (though availability is expected to be very limited).

When two or more agents are being evaluated in the same phase of the study, the trial design includes sharing of the appropriate control group (placebo or active comparator) for efficient evaluation of each agent. Note, however, that enrollment to the phase III placebo-controlled evaluation of BRII-196+BRII-198 did not coincide with enrollment the phase III placebo-controlled evaluation of SAB-185 and so there is no sharing of the placebo control group for these two agents.

Eligible participants enrolled under all versions of the protocol from version 2.0 have intensive follow-up through day 28, followed by limited follow up through at least week 72 weeks in phase II and phase III.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

Under protocol version 8.0, the phase II trial and the phase III trial involve different populations and have separate randomizations. However, the structure of the randomization process is the same for each of the two trials, as described in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of concurrently randomized participants on a given investigational agent and in the placebo comparator group for that agent (i.e. combining participants who were eligible to receive the agent but who were randomized to any of the available placebos for investigational agents in the same phase of evaluation).

To achieve having a similar number of participants on the active arm and in the pooled comparator group for a given investigational agent, the randomization occurs in two steps within each trial.

The first randomization is to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents in the trial (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). Trial phase for an agent is accounted for in the participant eligibility (i.e. by the classification of their risk for hospitalization/death as lower or higher). In the event that a

participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

Immediately following the first randomization, participants are randomized within an Agent Group in the second randomization to the investigational agent or appropriate comparator (i.e., the matching placebo for agents in the same phase of evaluation). For a given participant, the probability of assignment to the investigational agent or placebo in the second randomization depends on the number of agents currently under investigation that the participant was eligible to receive, as phase II and phase III have distinct populations (phase II is restricted to those at lower risk of progression to hospitalization and death, and phase III is restricted to those at higher risk).

Both the first and second randomizations involve blocked stratified randomization. In phase II, both the first and second randomizations are stratified by time from symptom onset (≤5 days vs >5 days), however, in previous versions of the protocol, in which both 'higher' and 'lower' risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group ('higher' vs 'lower'). In the active-controlled phase III trial introduced in protocol version 7.0 and the placebo-controlled trial introduced in protocol version 8.0, both randomization steps are stratified by country. Under previous versions of the protocol for the placebo-controlled phase III trial evaluating BRII-196+BRII-198, both randomization steps were only stratified by time from symptom onset (≤ 5 days vs > 5 days). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 8.0; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed outside of this SAP. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed separately.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, and 14 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical

needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in NP swabs [Protocol Objective 1.2.3].

4) Phase III: To determine the efficacy of the investigational agent to increase the proportion of participants with NP SARS-CoV-2 RNA below the LLoQ at study day 3 [Protocol Objective 1.2.4]

5) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.5].

6) Phase II and III: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72 [Protocol Objective 1.2.7].

8) Phase III: To evaluate if the investigational agent reduces the time to sustained symptom resolution through study day 28 [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of hospitalization or death through study day 28, excluding hospitalizations that are determined to be unrelated to COVID-19 [Protocol Objective 1.2.9].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes and differences between investigational agent and control (placebo in phase II and active comparator in phase III) across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course in those hospitalized [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 8.0. The sample size for the placebo-controlled superiority phase III design was also the same under protocol versions 2.0 to 6.0 (it was originally defined in Appendix IV of protocol version 2.0 for the BRII-196+BRII-198 agent entered into the study under protocol version 2.0) and is currently detailed in Appendix V of protocol version 8.0 for the BRII-196+BRII-198 agent. The sample size for the non-inferiority phase III design that enrolled participants under protocol version 7.0 is detailed in section 10.4 of

protocol version 7.0. As noted above, enrollment to that non-inferiority trial was terminated early. The sample size for the placebo-controlled phase III trial introduced in protocol version 8.0 is detailed in section 10.4 of protocol version 8.0.

2.4.1 Phase II – Placebo-Controlled Superiority

For each investigational agent in phase II, the proposed sample size is 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

2.4.2 Phase III – Placebo-Controlled Superiority Trial Used for the Evaluation of BRII-196+BRII-198

The proposed sample size was 842 participants consisting of 421 participants who received the active agent and 421 participants who were concurrently randomized to placebo control. This sample size included the enrollment that occurred during the phase II placebo-controlled evaluation of an agent. Participants who were randomized but did not start their randomized investigational agent or placebo were not followed.

This sample size was chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This was based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.4.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant)

The active-controlled Phase III trial was focused on a non-inferiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus an active comparator agent, specifically the monoclonal antibody combination of casirivimab plus imdevimab. The non-inferiority margin for the absolute difference in proportion hospitalized/dead was 3% (investigational agent minus active comparator agent); the rationale for this choice was described in Section 3.1 of protocol version 7.0. Non-inferiority was considered to be established if a two-sided exact 95% confidence interval for the absolute difference was entirely below 3%. Details of the construction of the confidence interval are in section 10.6 of protocol version 7.0 and are included further below in this SAP.

The sample size differed between infused investigational agents (600 for the investigational agent and 600 for the concurrently randomized active comparator) and non-infused investigational agents (800 per arm instead of 600 per arm). The rationale for this was that there may be broader clinical utility for non-infused agents such that a slightly higher true hospitalization/death rate may be tolerated in clinical practice. No enrollment occurred for a non-infused agent and so the sample size justification described below is just for an infused agent (enrollment only occurred for the SAB-185 infused agent).

Sample Size Justification for Infused Investigational Agents

For the evaluation of a specific infused investigational agent, the sample size was 1200 participants including approximately 600 participants randomized to receive the infused investigational agent and approximately 600 participants (who were eligible to receive the infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size was chosen to provide close to 90% power to establish non-inferiority assuming that the true proportion hospitalized/dead for both the infused investigational agent and the active comparator agent was 2.3%. The rate of 2.3% was based on the observed proportion for casirivimab plus imdevimab combining across doses in the subpopulation of the Regeneron COV-2067 clinical trial who met the criteria for being at high risk of progression to hospitalization/death (FDA communication to DAIDS/NIAID, May 2021). No adjustment for loss to follow-up was made in the sample size as the primary analysis was to be based on the observed number of hospitalizations divided by the number of participants who initiated study treatment. In addition, the impact of any loss to follow-up was expected to be minimal as there was to be regular contact between research site staff and participants (or their secondary contacts) and previous experience in the study and other trials has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up).

The potential power of the study was evaluated in two ways using the PASS version 15 sample size calculation software. Both used a non-inferiority hypothesis testing approach based on use of the Miettinen and Nurminen score test statistic (which was the basis for calculating the confidence interval used for the analysis). The first ignored interim monitoring but used a binomial

enumeration method to calculate power and type I error rates. Use of the binomial enumeration method takes account of the discreteness of the binomial distribution (rather than using a normal approximation to the binomial distribution) which may be important in the setting of low hospitalization/death probabilities. Using this approach gave a power of 90.2%. The second approach did not use a binomial enumeration but took account of interim analyses using a standard implementation of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 90.0% (width of 95% confidence interval around this simulation-based value was 0.12%). Based on these two approaches, it was anticipated that the study would have had close to 90% power to show non-inferiority for an infused investigational agent assuming that it truly had the same 2.3% hospitalization/death rate as the active comparator agent.

The PASS software was also used to illustrate how the power of the study might change for various scenarios which differed from the scenario assumed (see Table below). This was undertaken using the first of the two approaches mentioned above (i.e., using the binomial enumeration approach). Looking at the top part of the table in which both the infused investigational agent and the active comparator agent have the same underlying true hospitalization/ death rate, the power is decreased if the true rate was above the assumed 2.3%, but increased if the true rate was less than 2.3%. If the true rate was 3%, then the power was still above 80%, but if the true rate is 4% it was reduced to 73%.

The middle and lower parts of the table show scenarios in which the infused investigational agent had a true hospitalization/death rate of 0.5% or 1% worse than the active comparator agent, respectively. If the true rate for the active comparator agent was 2.3% and was 2.8% for the infused investigational agent (i.e., 0.5% worse), then the power was reduced to 73%. If the true rate for the active comparator agent was 2.3% and was 3.3% for the infused investigational agent (i.e., 1% worse), then the power was reduced to 50%

Table: Power for various scenarios based on non-inferiority hypothesis testing using the likelihood score test statistic (Miettinen and Nurminen method) with binomial enumeration of power and Type I error rate. All scenarios use a 3% non-inferiority margin and one-sided Type-I error rate of 0.025 with a sample size of 600 participants receiving an infused investigational agent and 600 participants receiving the active comparator agent. Power in practice would have been slightly reduced from the values shown due to interim monitoring.

Same Underlying True Hospitalization/Death Rate For Active Comparator Agent and Infused Investigational Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
99.4%	1%	1%	2.2%
97.3%	1.5%	1.5%	2.2%
93.2%	2%	2%	2.3%
90.2%	2.3%	2.3%	2.4%
88.1%	2.5%	2.5%	2.4%
83.1%	3%	3%	2.4%
78.1%	3.5%	3.5%	2.4%
73.2%	4%	4%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 0.5% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
92.5%	1%	1.5%	2.2%
85.2%	1.5%	2%	2.2%
77.4%	2%	2.5%	2.3%
73.1%	2.3%	2.8%	2.4%
70.5%	2.5%	3%	2.4%
64.8%	3%	3.5%	2.4%
59.6%	3.5%	4%	2.4%
55.0%	4%	4.5%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 1% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
71.3%	1%	2%	2.2%
61.7%	1.5%	2.5%	2.2%
54.0%	2%	3%	2.3%
50.4%	2.3%	3.3%	2.4%
48.4%	2.5%	3.5%	2.4%
44.0%	3%	4%	2.4%
40.0%	3.5%	4.5%	2.4%
36.7%	4%	5%	2.4%
*Actual type I error rate is slightly lower than assumed rate of 2.5% because of discreteness of the binomial distribution.			

2.4.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0

The proposed sample size is 1200 participants consisting of approximately 600 participants who are randomized to receive the active agent and approximately 600 participants who are concurrently randomized to placebo control. Unlike the placebo-controlled phase III evaluation of

investigational agents under earlier versions of the protocol, under protocol version 8.0, participants enrolled in the phase II evaluation of an investigational agent are not part of the study population for the phase III evaluation of the same agent (as participants in the phase II evaluation were generally “lower risk” and participants in the phase III evaluation are “higher risk” for hospitalization/death). Participants who are randomized but do not start their randomized investigational agent or placebo are not followed.

The phase III trial is focused on a superiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus placebo (with use of SOC treatment in both arms, if available). The primary analysis will focus on evaluating the ratio of proportions (investigational agent/placebo) or, equivalently, the relative reduction in risk of hospitalization/death for the active investigational agent versus placebo. The sample size of 1200 participants, with approximately 600 randomized to an investigational agent and 600 to placebo, has been chosen to give good power (>90%) to detect relative risk reductions of 70% (as found for other antibody treatments) if the proportion hospitalized/dead in the placebo group is about 5% or higher, using a two-sided Type I error rate of 5%. There are multiple factors that will affect the power, which are discussed below. To provide context for this discussion, the table below shows the power of the study to detect relative risk reductions of between 50% and 70% for proportions hospitalized/dead in the placebo group of 3% to 6%. The powers shown were obtained in PASS software (version 15.0.4) for testing two proportions using a z-test (so the normal approximation method) with unpooled variance. They are based on an effective sample size of 570 per arm, with the 5% reduction from 600 per arm built in to allow for loss to follow-up and interim monitoring using the O’Brien and Fleming stopping guideline.

Power to detect various true effect sizes (relative reduction in risk of hospitalization/death) for selected true proportions hospitalized/dead on placebo between 3% and 6%			
Proportion Hospitalized/Dead		Relative Risk Reduction for Active versus Placebo	Power
Placebo	Active		
3%	0.9%	70%	73%
	1.2%	60%	56%
	1.5%	50%	40%
4%	1.2%	70%	85%
	1.6%	60%	69%
	2.0%	50%	51%
5%	1.5%	70%	92%
	2.0%	60%	79%

	2.5%	50%	61%
6%	1.8%	70%	96%
	2.4%	60%	86%
	3.0%	50%	69%

Discussion of Factors Affecting the Power of the Study

a. Proportion hospitalized/dead in placebo control group: As can be seen in the table, the proportion hospitalized/dead in the placebo arm has a reasonable effect on power with lower proportions leading to a reduction in power. For the placebo control group for evaluating the BRII agent in ACTIV-2, the proportion was 11% [25]. However, the proportion in the phase 3 trial of sotrovimab was 6% [26]. There is also a possibility that the proportion may be lower for the Omicron variant than with previous variants.

b. Use of SOC treatment by some participants: Higher use of SOC treatment will reduce the proportion hospitalized/dead in both randomized arms. For example, the proportion hospitalized/dead in the placebo arm would change from 6% if none receive SOC treatment to 5.58%, 4.74% and 3.90% if SOC treatment is used by a random sample of 10%, 30% and 50%, respectively, of participants in the placebo arm (i.e., SOC treatment use is not related to risk of hospitalization/death) and SOC treatment reduces risk of hospitalization/death by 70%. If SOC treatment use is not random, for example, it is taken up by the highest risk participants, then the impact might be larger. As the trial excludes participants who have accessed SOC treatment prior to entry and there is a general lack of availability of such treatments globally, use in the trial is expected to be very low (e.g., <10%) and so will limit the impact.

c. Differential effect of an investigational agent versus placebo according to use or not of SOC treatment: For a given proportion of participants hospitalized/dead in the placebo arm, the power shown in the above table is valid if the relative effect of SAB versus placebo is not affected by the use of SOC treatment. Power would be reduced from the values shown if the effect of SAB versus placebo is reduced in the presence versus absence of background therapy. A related concern arises if use of SOC treatment is differential in the investigational agent arm versus the placebo arm. For example, accessing SOC treatment at a higher rate in the placebo arm because more participants have a deteriorating health status might diminish a true difference in effect between arms and hence reduce power. As noted above, use of SOC treatment is expected to be low and so any reduction in power is expected to be limited even if this occurs.

d. Failure to start randomized treatment and loss to follow-up: The impact of any loss to follow-up is expected to be minimal as there will be regular contact between research site staff and participants (or their secondary contacts). Previous experience in the study and other trials

has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up) when loss to follow-up has also been minimal (approximately 1 to 1.5%) in this study. In addition, a very small proportion of randomized participants will not start study treatment and will be excluded from the analysis of the primary outcome. Based on ACTIV-2 experience, allowance for 3-4% not starting treatment or being lost to follow-up before hospitalization is built into the above power table (with additional allowance of 1-2% for interim monitoring using the O'Brien and Fleming stopping guideline).

Because of these uncertainties, the DSMB will be asked to monitor the potential impact of the above factors on the operational feasibility of the study.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring of the placebo-controlled superiority phase II, the placebo-controlled superiority phase III for evaluating BRII-196+BRII-198, the active-controlled non-inferiority phase III for evaluating SAB-185 under protocol version 7.0, and the placebo-controlled superiority phase III for monitoring investigational agents under protocol version 8.0. Additional details on phase II monitoring can be found in protocol version 8.0 section 10.5.1, and in protocol version 8.0 Appendix V for placebo-controlled phase III monitoring for the BRII-196+BRII-198 agent. Details on active-controlled non-inferiority phase III monitoring for SAB-185 are taken from protocol version 7.0 section 10.5.2 as amended in letter of amendment 1 to protocol version 7.0. Details on the placebo-controlled superiority phase III monitoring introduced under protocol version 8.0 are taken from section 10.5.2 of protocol version 8.0. Statistical considerations for interim monitoring are described in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to study product or if two participants experience a Grade 4 AE deemed related to study product, enrollment to the study product group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II – Placebo-Controlled Superiority

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III – Placebo-Controlled Superiority Used for the Evaluation of BRII-196+BRII-198

[At the time of preparing this version of the SAP, all participants randomized to receive BRII-196+BRII-198 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study].

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

2.5.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant)

[At the time of preparing this version of the SAP, all participants randomized in the non-inferiority phase III evaluation of SAB-185 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study. Note that the first interim analysis for this part of the study was scheduled before the termination of enrollment due to the anticipated lack of efficacy on the control regimen against the SARS-CoV-2 Omicron variant that had become widely prevalent. That interim analysis was replaced by an interim analysis that followed the monitoring approach described in Section 2.5.4 and the SAB-185-specific appendix of this SAP].

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if there is persuasive evidence of non-inferiority (or superiority or inferiority) of an investigational agent versus the active comparator agent in its effect on the hospitalization/death outcome. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend

termination or modification of the study if it appears futile on statistical or operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, three interim analyses for DSMB review are planned for each investigational agent, after approximately 25%, 50% and 75% of the planned enrollment for an investigational agent has been completed and followed through to day 28. At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guidelines for Efficacy or Lack of Efficacy

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses) for the absolute difference (investigational agent minus active comparator agent) in the proportion of participants hospitalized or dead by day 28, relative to thresholds defining non-inferiority, superiority or inferiority of the investigational agent as follows (in the order given):

- The DSMB may recommend releasing results evaluating the effect of an investigational agent when both non-inferiority and superiority of that agent is established based on the confidence interval being entirely below 0% (i.e., supportive of a lower true proportion being hospitalized or dying on the investigational agent than the active comparator agent). If this occurs, consideration will need to be given to the ongoing appropriateness of the active comparator agent as a control for evaluating other investigational agents in the study.

- Early stopping and/or release of results based on non-inferiority should be considered on an agent-by-agent basis. For non-infused agents, the DSMB may recommend releasing results evaluating the effect of an investigational agent when non-inferiority (but not superiority) of that agent is established based on the confidence interval being entirely below 3% (but not entirely below 0%). However, in the interests of also having an adequate safety database for the investigational agent, it is not intended that this recommendation be made before approximately 400 participants have been randomized to receive the agent (or some other number of participants specified in the agent-specific appendix). In addition, the study may continue randomizing participants to the investigational agent in the interests of increasing precision in evaluating the agent; this decision will be made by the study team and sponsor on an agent-by-agent basis. For infused agents, early stopping and/or release of results for non-inferiority should not be considered.

- The DSMB may recommend releasing results and terminating randomization to an investigational agent if inferiority of that agent is established based on the confidence interval being entirely above 0% (i.e., suggesting a higher true proportion being hospitalized or dying on the investigational agent than the active comparator agent). Examples of how this criterion might be met when evaluating an infused agent and when

the observed control rate is close to 2.3% include observing 18/150 versus 3/150 (observed difference 10.0%) at the first interim analysis; 23/300 versus 7/300 (observed difference 5.3%) at the second interim analysis; and 24/450 versus 10/450 at the third interim analysis (observed difference 3.1%). In these examples, all observed differences are higher than the non-inferiority margin of 3%, and are indicative also of the futility of continuing evaluation of the infused investigational agent to demonstrate non-inferiority.

2.5.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if efficacy of the agent versus placebo has been established, or if it is unlikely that the agent has sufficient efficacy to warrant further evaluation in this study. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend termination or modification of the study if it appears futile on operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, two interim analyses for DSMB review are planned for each investigational agent, after approximately one-third (i.e., approximately 400 participants) and two-thirds (i.e., approximately 800 participants) of the planned enrollment for an investigational agent has been completed and followed through to day 14 (the choice of day 14 is because the large majority of hospitalizations/deaths in ACTIV-2 have been observed to occur by day 14).

At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guideline for Efficacy Favoring an Investigational Agent versus Placebo

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses—see further below) for the relative difference (investigational agent / placebo) in the proportion of participants hospitalized or dead by day 28. As a stopping guideline for greater efficacy of an investigational agent compared with placebo, the O'Brien and Fleming boundary will be used. The stopping guideline will be implemented using the Lan-DeMets spending function approach to allow for the possibility of changes in the timing of interim analyses and/or additional (or fewer) interim analyses if recommended by the DSMB. Information time for the spending function will be based on the proportion of the planned enrollment (i.e., of the 1200 participants for comparing an investigational agent to placebo) who could have been followed through day 14 at the time of the data freeze for the interim analysis. The choice of day 14 here reflects the fact that the very large majority of hospitalizations and deaths in ACTIV-2 have occurred by 14 days of follow-up. As a guideline, if the two-sided 95% confidence interval (adjusted for interim analyses) excludes a risk

ratio of one (equivalently a relative risk reduction of zero) favoring the investigational agent, then the DSMB may recommend closure of randomization to that agent; release of interim results may also be recommended.

There is the possibility that differences between the treatment groups may be observed early in follow-up. However, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the treatment group comparisons will be on the cumulative proportion hospitalized/dead at day 28.

Stopping Randomization to an Investigational Agent Because of Limited Efficacy

Because there are treatments available that may substantially reduce the risk of hospitalization/death, albeit with limited availability and the caveat that they have generally been evaluated among individuals infected with earlier variants of SARS-CoV-2, it is likely that a treatment which reduces the risk of hospitalization/death by less than 30% versus placebo will have limited utility in clinical practice. Therefore, as a non-binding guideline, the DSMB may recommend early termination of randomization to a specific investigational agent because of limited efficacy if the two-sided 95% confidence interval (adjusted for interim analyses) for the risk ratio is entirely above 0.7 or, equivalently, the two-sided 95% confidence interval (adjusted for interim analyses) for the relative risk reduction is entirely below 30%.

Modifying or Stopping the Study for Operational Futility

The DSMB will also monitor operational futility, in particular related to losses to follow up, low hospitalization/death rate in the placebo arm (which, in part, may arise due to more extensive use of SOC treatment than anticipated). As most hospitalizations are expected to occur early in follow up (e.g., during the first 14 days), early losses to follow up would be most relevant. As a benchmark, an overall loss to follow-up rate (excluding losses after a participant is hospitalized) of more than 5% would be cause for concern.

With regard to the hospitalization/death rate in the placebo arm, the power of the study is limited if this rate is below 3% (see power analysis table above). Therefore, as a benchmark, an observed rate of less than 3% in the placebo arm would be a cause for concern. If this arises, or temporal trends in hospitalization/death rate suggest it might, then any DSMB recommendation concerning this issue might incorporate information about factors that might be driving it (e.g., increasing use of SOC treatment, evolving lower risk of participants enrolled, or lower risk with new variants).

2.6 Graduation to Phase III

[At the time of preparing this version of the SAP, all agents that had been initiated in phase II evaluation have been evaluated for graduation to phase III evaluation or a decision has been made not to evaluate them for graduation. The following therefore provides a summary of the approach to graduation that is in protocol version 8.0 recognizing that there are currently no further agents being considered for graduation].

Each investigational agent that is being considered for evaluation in phase III will be evaluated for safety, for activity in reducing COVID-19 symptoms and hospitalization/death, and for activity in reducing SARS-CoV-2 RNA shedding. An analysis to determine if an agent should graduate from phase II, and enter phase III, will be conducted when 220 participants assigned to the agent or concurrent placebo in phase II evaluation have completed their Day 7 evaluations and have the required data available in the database. Additional interim graduations may be assessed for some agents, see agent-specific appendices in the protocol for details.

The DSMB will review unblinded data and make recommendations to NIAID (as trial sponsor) and to the TOC, indicating whether graduation criteria have been met. The recommendation for an agent to enter phase III evaluation will be made by the TOC in discussion with the collaborating company; the collaborating company that is responsible for the agent will decide whether or not to adopt the recommendation. The TOC and collaborating company will also consider which dose to recommend for evaluation in phase III, for investigational agents with more than one dose under evaluation in phase II. NIAID/DAIDS, as the sponsor of the study, will make the final determination regarding graduation of the study product.

The TOC may recommend an agent move directly into phase III, without evaluation in phase II in ACTIV-2, if there is sufficient safety and efficacy data supporting phase III evaluation available from outside of the trial. These agents will not undergo graduation analyses.

Graduation criteria and statistical considerations are discussed in the Graduation Rules SAP.

3 Outcome Measures

All outcome measures are copied from the protocol version 8.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo [active comparator intervention under protocol version 7.0]. [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the number of days from start of investigational treatment to the first of two consecutive days when all symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: At each of days 3, 7 and 14 quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs. [For Primary Objective 3]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
   [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at

baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment. [Supportive of Primary Objective 2]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment. [Supportive of Primary Objective 2 and for Secondary Objective 8]

   Duration defined as the number of days from start of investigational treatment to the first of four consecutive days when all symptoms are scored as absent. Targeted symptoms as defined in the primary phase II clinical (symptom duration) outcome.

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28. [Supportive of Primary Objective 2]

   Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of two consecutive days that a participant reported return to usual (pre-COVID) health.

7) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28. [Supportive of Primary Objective 2]

   Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

8) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention, hospitalization, and death. [For Secondary Objective 1].

   Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

9) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or comparator intervention. [For Secondary Objective 2]

10) Phase II and Phase III: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

11) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

12) Phase III (Active-Controlled [protocol version 7.0] and placebo-controlled [protocol version 8.0]) only: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs at day 3. [Support of Primary Objective 3]

13) Phase II and III: Level (quantitative) of SARS-CoV-2 RNA from staff-collected NP swabs at days 3, 7, and 14 in phase II and at day 3 in phase III.8.
[For Secondary Objective 3]

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from staff-collected NP swabs at days 0, 3, 7, and 14. [Supportive of both Primary Objective 3 and Secondary Objective 3]

Efficacy

15) Phase II only: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

17) Phase II and III: Death due to any cause or hospitalization due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death due to any cause or hospitalization due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

19) Phase II and III: Death due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

20) Phase II and III: Death due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

21) Phase III: Death due to any cause or hospitalization due to any cause, excluding hospitalizations that are deemed unrelated to COVID-19, during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 9]

    Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic. An Adjudication committee is evaluating the relatedness of hospitalization due to COVID-19.

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or comparator intervention. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

    Ordinal scale defined as:

    Death
    Hospitalized, on invasive mechanical ventilation or ECMO;
    Hospitalized, on non-invasive ventilation or high flow oxygen devices;
    Hospitalized, requiring supplemental oxygen;
    Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

   Ordinal scale defined as:
   - Death
   - Hospitalized, on invasive mechanical ventilation or ECMO;
   - Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   - Hospitalized, requiring supplemental oxygen;
   - Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to its comparator intervention (placebo or active comparator, as appropriate for the agent and phase of evaluation).

- Treated Population: All participants in the Randomized Population who received any investigational agent or its comparator agent (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus its comparator intervention. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to the comparator intervention, who were also eligible to have received the investigational agent of interest. For the placebo-controlled trials, the comparison group will pool across all relevant placebos (i.e. including the placebo for the agent of interest and the placebos for other agents). For the primary placebo-controlled analysis of a specific investigational agent, a supplemental analysis may be undertaken that restricts the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/comparator intervention occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/comparator intervention.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24).

Week 72: Key visit for evaluation longer-term efficacy and safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLoQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

Virology results generated from specimens with the following conditions reported in the database will be excluded from analyses:

- Thawed;
- Invalid Specimen;
- Quantity Not Sufficient;
- Destroyed.

Note: Samples with the condition code ‘NOT’ were also to be excluded per the trial sponsor but this code indicates that the specimen was not tested. Thus, no result is expected and no exclusion is needed.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or comparator intervention are not followed.

Participants who have protocol violations, such as those who start investigational agent or comparator intervention outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at “higher” risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the placebo-controlled superiority design. Further details are provided after the table.

<table>
<tr><td colspan="3">Phase III Primary Objective for Efficacy—Placebo-Controlled Superiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.</td></tr>
<tr><td>Estimand description</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td colspan="2">Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td colspan="2">Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td colspan="2">Treated Population</td></tr>
<tr><td>Variable(s)</td><td colspan="2">Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td colspan="2">Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td colspan="2">Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td colspan="2">Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td colspan="2">Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td colspan="2">Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="3">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5), and to 7 days under protocol version 6 and subsequent protocol versions.</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (due to any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17, 18, 19, 20 and 21, evaluate the proportion of participants who are hospitalized or died through week 24, the proportion who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days) or through 504 days (i.e. 72 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher’s exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the active-controlled non-inferiority design. Further details are provided after the table.

Phase III Primary Objective for Efficacy—Active-Controlled Non-Inferiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.	
Estimand description	Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.
Treatment	Investigational agent or active comparator agent (casirivimab and imdevimab).

Target population	Analysis set (analysis population)
Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry	Treated Population
Variable(s)	**Outcome measure(s)**
Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent (coded as 1 if participant died or was hospitalized, and 0 otherwise).	Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent.
Handling of intercurrent events	**Handling of missing data**
None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of the agent to which they were randomized.	Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as not having an event after the date last known to be alive.
Population-level summary measure	**Analysis approach**
Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization over 28 days.	Difference (for investigational agent minus active comparator agent) of the proportion dying or being hospitalized at day 28. See text for further details.

Analysis Approach

The analysis of the primary efficacy outcome in phase III will evaluate the absolute difference in proportion of participants hospitalized (due to any cause) or died (due to any cause), from day 0 through day 28, between randomized arms; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included.

Inference will be based on constructing a two-sided exact 95% confidence interval for the absolute difference in proportions (proportion for the investigational agent minus the proportion for the active comparator agent). If this confidence interval is entirely below the non-inferiority margin of 3%, then a conclusion of non-inferiority of the investigational agent compared with the active

comparator agent will provide reasonable evidence that the investigational agent is effective against COVID-19.

The exact 95% confidence interval will be calculated using the method of Chan and Zhang [Biometrics 1999;55:1201-09] as implemented, for example, in StatXact PROC BINOMIAL for SAS [StatXact 12 PROCs for SAS Users Manual. Cytel Inc., Cambridge, MA; 2019]. This method inverts two one-sided hypothesis tests (with one-sided error rate of 0.025 each) to obtain the confidence interval so providing a confidence interval-based method which preserves the type I error rate in establishing non-inferiority to be 0.025. To preserve confidence interval coverage (and type I error rate for assessing non-inferiority) over multiple interim analyses, the confidence interval will be calculated using a “repeated” confidence interval approach with spending of error rate at each interim analysis using the Land and DeMets approach with an O’Brien and Fleming spending function.

In essence, basing the comparison of treatment groups on the simple proportion of participants who were hospitalized or died assumes that participants who are lost to follow-up before 28 days without prior hospitalization were not hospitalized and did not die by 28 days. The decision to use the simple proportion for analysis rather than use, for example, a Kaplan-Meier estimate of the cumulative proportion of participants hospitalized or dying during the first 28 days of follow-up to account for losses to follow-up was taken for multiple reasons. First, in ACTIV-2 and other COVID-19 trials, most hospitalizations and deaths occur during the first two weeks of follow-up and the study has been designed to have regular contact with participants or their secondary contacts so as to maximize ascertainment of hospitalization and death information. Second, loss to follow-up has been low in the ACTIV-2 study: approximately 3% among higher risk participants. Third, with the very low rates of hospitalization/death expected (e.g., 2.3% for the active comparator agent), confidence interval coverage (and type I error rates) are better preserved at their desired levels through the use of exact statistical methods for analyzing proportions than is achieved using asymptotic statistical methods based on Wald-type analyses using Greenwood’s formula to obtain standard errors for Kaplan-Meier estimates. To assess the potential impact of loss to follow-up (assumed to be non-informative) on the interpretation of results, the following sensitivity analyses will be undertaken, repeating the primary analysis repeated with:

(a) a comparison of the simple proportions using a Wald-based confidence interval; and

(b) a comparison of proportions estimated using Kaplan-Meier methods (with censoring of follow-up at the earlier of day 28 and the time that a participant was last known to be alive) using a Wald-based confidence interval with standard error based on Greenwood’s formula.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 16, 17, 18, 19, 20 and 21 evaluate the proportion of participants who die through to day 28, the proportion who are hospitalized or died through week 24, the proportion

who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In the sensitivity analyses based on Kaplan-Meier estimates, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days for outcomes through to 24 weeks) or through 504 days (i.e. 72 times 7 days for outcomes through to 72 weeks).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approach outlined for the primary analysis will be implemented for each subgroup. However, these analyses are likely to involve small numbers of events in most or all subgroups and hence have very limited precision. Because of this, any assessment of treatment by subgroup interaction, if undertaken, will be considered exploratory. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/active comparator Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or comparator intervention) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

In placebo-controlled evaluations, because some agents may be administered using injections or infusions and others will not be, the primary safety analysis may be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed in the same manner as the primary safety outcomes.

Additional longer-term safety outcomes may be assessed, see agent-specific appendices for details.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the number of days from start of investigational agent (day 0, pre-treatment) to the first of two consecutive days when all symptoms scored as moderate

or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and comparator intervention arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic

missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to “absent”, and (b) four consecutive days with resolution of all targeted symptoms to “absent” (i.e., secondary outcome measure 5). For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are “absent” on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis maybe repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to comparator intervention to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) ‘Risk of Severe Disease’ Stratification [this may not be pursued for agents which predominantly enrolled participants who were at ‘lower’ risk for severe COVID progression]
5) Age Group (<60, ≥60)

6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, and 14).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as "0" if above LLoQ, and "1" if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or comparator intervention) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the number of days from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;
3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model

may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

Duration of Symptoms

In support of the symptom duration outcome in phase III, the analysis will be repeated using the same approach described in the primary symptom duration analysis for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to “absent.” To address secondary objective 8, and in supportive of the symptom duration outcome in phase III, a similar TTE outcome measure will also be examined defined as time to four consecutive days with resolution of all targeted symptoms to “absent,” (i.e. secondary outcome measure 5).

Return to Usual Health

The analysis of return to usual health will be repeated using the same approach described above for a similar TTE outcome measures defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2.2 Secondary Virology

The schedule of evaluations in protocol version 7.0 indicates that only NP swabs will collected in both phase II and phase III, and therefore only analyses of SARS-CoV-2 RNA from NP swabs are outlined below. Some agents may have completed enrollment in phase II prior to implementing protocol version 7.0, and therefore may have additional specimens collected for SARS-CoV-2 RNA testing. If analyses of these additional specimens are pursued, then the approach will be as defined in the relevant previous version of the SAP.

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA at day 3 (this is a secondary outcome for the active-controlled phase 3 only)

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from staff-collected NP swabs at entry and day 3.

The proportion of participants with SARS-CoV-2 RNA < LLoQ day 3 will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. The model will include a main effect for treatment and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained by taking the exponential of the treatment parameter estimate (and associated 95% CI). In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ). Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time for staff-collected NP swabs.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

In phase II only, levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs will be analyzed using participant-specific AUCs. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and

above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: max[0, $\log_{10}$(RNA)-$\log_{10}$(LLoQ)], obtained at the scheduled measurement times between and including day 0 and day 14.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 14 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 14 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent comparator intervention through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent comparator intervention through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent comparator intervention who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with

frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 72.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

Interim analyses of the placebo-controlled superiority phase III evaluation of an agent was finished at the time of finalization of SAP version 7.0. The following from protocol version 7.0 describes the interim analysis considerations for the active-controlled non-inferiority phase III evaluation of an agent.

The two-sided 95% confidence interval mentioned above [see section 2.5.3 of the SAP] will be adjusted for the multiple interim analyses to preserve the confidence interval coverage to at least 95% (this is also referred to as using “repeated” confidence intervals).

The standard Lan and DeMets approach will be used to achieve this, incorporating an O’Brien and Fleming spending function. For simplicity, the information scale for the spending function will be determined as the proportion of the planned enrollment randomized to the investigational agent being evaluated at the time of the interim analysis. As an example, if in practice, the analyses were after exactly 25%, 50%, 75% and 100% of the planned enrollment, then the nominal confidence intervals used to assess efficacy would have coverage 99.9985% at the first analysis, 99.70% at the second analysis, 98.17% at the third analysis and 95.60% at the fourth analysis (these were obtained from PASS software). However, as the O’Brien and Fleming spending function is very conservative at early interim analyses, making stopping very difficult, for the assessment of inferiority of an investigational agent compared to the active comparator agent, an asymmetric approach will be used to reduce the level of evidence required for early stopping in the event that an investigational agent appears inferior to the active comparator agent. Specifically, if a nominal confidence interval with coverage of greater 99.9% at an early interim analysis is suggested by use of the O’Brien and Fleming spending function, then a nominal confidence interval with coverage of 99.9% will be used instead for assessing inferiority of the investigational agent.

The DSMB will also monitor the proportion hospitalized/dead in the active comparator arm as this key parameter, coupled with the non-inferiority margin, underpins the study design. The study is designed assuming that the underlying true proportion of participants on the active comparator agent is 2.3%. This is the proportion (32/1392) observed for high risk participants in the Regeneron COV-2067 trial for the agent (pooling across doses studied in that trial; FDA communication to DAIDS/NIAID). A 95% confidence interval for this proportion is (1.5%, 3.1%).

An assessment of non-inferiority in this study would be more difficult if the proportion of participants on the active comparator agent in this study is somewhat different from that in the Regeneron COV-2067 (e.g., somewhat outside of the range suggested by the confidence interval).

For example, this might arise if variants of SARS-CoV-2 are present in the study population which the active comparator agent is less effective against. Such an issue would undermine the use of a 3% non-inferiority margin in this study. It may however be addressed by focusing the non-inferiority assessment on the subpopulation in this study without such variants (assuming these have been identified), or in establishing superiority of the investigational agent in the overall study population. This may require a larger sample size to maintain power.

Another potential reason for a somewhat different proportion hospitalized/dead on the active comparator agent in this study versus that in the Regeneron COV-2067 study is that this study is likely to enroll in a number of different countries, whereas the Regeneron COV-2067 enrolled primarily in the United States. Aside from possible differences in circulating variants among countries, differences among countries in clinical practice and/or in the availability of hospital care might lead to differences in hospitalization/death rates. The DSMB will monitor descriptive results by country and provide guidance about countries with notably low or high rates of hospitalization/death.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during

hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued through to protocol version 6.0. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions. All participants enrolled to evaluate BRII-198+BRII-196 were randomized to active agent or placebo and so placebo is mentioned as the comparator intervention throughout this appendix.

7.1 Randomization Details

Phase III is stratified by time from symptom onset (≤ 5 days versus > 5 days).

7.2 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.3 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 72).

8 Appendix 3: Statistical Considerations for AZD7442 IV

NOTE: AZD7442 IV is only being evaluated in this study in phase II with a placebo control.

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 72).

9 Appendix 4: Statistical Considerations for AZD7742 IM

NOTE: AZD7442 IM is only being evaluated in this study in phase II with a placebo control.

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 72).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

2) Phase III: To determine whether SNG001 reduces hospitalization or death through study day 28 among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

3) Phase III: To determine whether SNG001 reduces duration of targeted COVID-19-associated symptoms through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

4) Phase III: To determine whether SNG001 reduces COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

5) Phase III: To determine whether SNG001 reduces progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at entry, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

6) Phase III: To determine whether SNG001 increases proportion of individuals with pulse oximetry measurement of ≥ 96% through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

7) Phase III: To determine whether SNG001 reduces the time to sustained symptom resolution through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

8) Phase III: To determine whether SNG001 prevents the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72, among individuals in the subgroup who report moderate or severe shortness of breath or

difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

9) Phase III: To determine whether SNG001 prevents the composite endpoint of hospitalization or death through stay day 28, excluding hospitalizations that are determined to be unrelated to COVID-19, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

10.1.2 Exploratory Objectives

1) Phase II and III: To determine whether SNG001 reduces severity of shortness of breath or difficulty breathing through study day 28.

2) Phase II: To determine whether SNG001 reduces a COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 among individuals who report moderate to severe shortness of breath or difficulty breathing at day 0.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II: Area under the curve of *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *shortness of breath or difficulty breathing* symptom in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

10.3.1 Secondary Analyses

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in

a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

The secondary objectives addressing analyses among people who reported moderate or severe shortness of breath or difficulty breathing at day 0, and among people who reported severe shortness of breath or difficulty breathing at day 0, will be undertaken in the same manner as the analyses of this outcomes among the overall study population.

10.3.2 Exploratory Analyses

Exploratory analyses will compare the AUC for cough and shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate.

To address SNG001-specific exploratory objective 2, the COVID-19 severity ranking outcome will compared between arms using the same methods outlined for the secondary analysis of this outcome measure, but restricted to be among those with moderate to severe shortness of breath or difficult breathing at day 0,

11 Appendix 6: Statistical Considerations for Camostat

Note: camostat is only being evaluated in this study in phase II with a placebo control.

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration in phase II (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a separate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

However, there are key special study design and analysis considerations for the phase III evaluation of SAB-185 (3,840 Units/kg). These are described in this appendix based on material presented in sections 3.4 and 10 of protocol version 8.0.

12.1 Study Design and Analysis Population Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185

Phase III evaluation of SAB-185 was initiated under protocol version 7.0 in a non-inferiority comparison of SAB-185 to an active control of casirivimab plus imdevimab. While enrollment was ongoing, the Omicron variant of SAR-CoV-2 became highly prevalent. In vitro data suggested that casirivimab plus imdevimab would be ineffective against this variant, and FDA authorization for emergency use of this regimen for treatment of COVID-19 was withdrawn due to non-susceptible SARS-CoV-2 variants (such as Omicron). Because of this, enrollment into the study was paused pending development of protocol version 8.0, which replaces the non-inferiority evaluation of investigational agents compared to casirivimab plus imdevimab with a placebo-controlled superiority design allowing for the additional use of standard of care (SOC) treatments, if available, in both arms.

Over 700 participants were enrolled under protocol version 7.0 and randomized to SAB-185 or casirivimab plus imdevimab. These participants can be divided into two subpopulations: (1) participants who were definitely or very likely infected with the Omicron SARS-CoV-2 variant (“Omicron Subpopulation”); and (2) participants who were definitely not, or likely not infected with the Omicron variant (“Non-Omicron Subpopulation”). Following the details in section 10.1.1 of protocol version 8.0, these two subpopulations are defined in more detail:

- The “Omicron Subpopulation” enrolled under protocol version 7.0 is defined as (1) all participants randomized under protocol version 7.0 who were infected with the Omicron variant as identified on sequencing of an NP sample obtained on day 0, plus (2) all participants randomized under protocol version 7.0 on or after December 26, 2021, who do not have variant information available from a sample obtained on day 0. The second of these two groups of participants are assumed very likely to be infected with the Omicron variant on the basis that prevalence of the Omicron variant in the U.S. was estimated by the CDC to be 89.2% for the week ending January 1, 2022 (and starting December 26, 2021), 95.2% for the week ending January 8, 2022, 97.9% for the week ending January 15, 2022, and 98.9% for the week ending January 22, 2022, during which enrollment under protocol version 7.0 was stopped [COVID Data Tracker. United States Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#variant-proportions. Accessed February 11, 2022].

- The “Non-Omicron Subpopulation” enrolled under protocol version 7.0 is defined as all participants enrolled under protocol version 7.0 excluding those in the “Omicron Subpopulation.” It therefore includes all participants randomized on or before December 25, 2021, who did not have the Omicron variant identified in sequencing of an NP sample obtained on day 0 (including those with no sequence available from an NP sample obtained on day 0). It also includes all participants randomized on or after December 26, 2021 who had a non-Omicron variant identified in sequencing of an NP sample obtained on day 0.

Based on in vitro data, the combination of casirivimab plus imdevimab is thought to have no effect on hospitalization/death in the Omicron Subpopulation and so is considered functionally to be a placebo from an efficacy perspective. Therefore, for the purposes of evaluating the superiority of SAB-185 versus placebo under this version of the protocol, the Randomized Population will be comprised of the Omicron Subpopulation enrolled under protocol version 7.0 as well as the population enrolled under protocol version 8.0 that is randomized to SAB-185 or its appropriate placebo control group. The planned sample size for this Randomized Population combining participants in the Omicron Subpopulation enrolled under protocol version 7.0 and participants enrolled under this version of the protocol is 1200 participants. The Treated Population (modified intent-to treat [mITT] population) for evaluating SAB-185 is this Randomized Population after excluding any participants who did not receive study product (i.e. excluding those who did not start SAB-185 or casirivimab plus imdevimab if enrolled under protocol version 7.0, and excluding those who did not start SAB-185 or placebo if enrolled under protocol version 8.0).

The Non-Omicron Subpopulation enrolled under protocol version 7.0 is not part of the phase III placebo-controlled superiority evaluation of SAB-185. See section 12.4 of this appendix for analysis considerations for this population.

12.2 Additional Analysis Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185

Data from the Treated Population defined in section 12.1 of this appendix will be used to evaluate the efficacy and safety of SAB-185 versus placebo (with possible SOC treatment added in both arms, if available). These analyses will follow the analysis plans in this SAP for the phase III placebo-controlled superiority evaluation of an investigational agent.

For the primary efficacy outcome measure (hospitalization/death), results comparing randomized arms may be presented separately for the Omicron Subpopulation enrolled under protocol version 7.0 and the population enrolled under protocol version 8.0. The possibility of heterogeneity in the effect of SAB-185 versus casirivimab plus imdevimab in the Omicron Subpopulation (considered functionally to be a placebo in this population) versus the effect of SAB-185 versus placebo (with the possibility of SOC treatment, if available) among participants enrolled under protocol version 8.0 may also be evaluated to assess the possible impact on interpretation of results. This analysis will follow the subgroup analysis plan for the primary efficacy outcome described in section 5.1.2 of this SAP.

For the major secondary outcome measures (4) symptom duration, (12) SARS-CoV-2 <LLoQ versus ≥LLoQ, and (13) quantitative SARS-CoV-2 RNA levels, results comparing randomized arms may also be presented separately for the Omicron Subpopulation enrolled under protocol version 7.0 and the population enrolled under protocol version 8.0. These analyses will follow the subgroup analysis plan for these secondary outcome measures described in section 5.2 of this SAP.

In addition, safety analyses will be presented separately for the following mutually exclusive subgroups:

(1) The Omicron Subpopulation enrolled under protocol version 7.0, as the control group received casirivimab plus imdevimab;

(2) Participants enrolled under protocol version 8.0, as the control group received placebo.

An additional breakdown of subgroup (2) will be undertaken for safety analyses as some participants may have received SOC treatment in addition to randomized SAB-185 or placebo:

(2a) Participants enrolled under protocol version 8.0 who did not receive SOC treatment; and

(2b) Participants enrolled under protocol version 8.0 who received SOC treatment.

It is recognized that the comparisons in subgroups (2a) and (2b) may not be pure randomized comparisons because receipt of SOC treatment may be influenced by the clinical status of a participant after randomization.

12.3 Analysis Considerations for the Non-Omicron Subpopulation Enrolled under Protocol Version 7.0

Follow-up of the Non-Omicron Subpopulation enrolled under protocol version 7.0 will continue per protocol. In this subpopulation, the combination of casirivimab plus imdevimab is expected to be effective. Analysis of outcomes from the Non-Omicron Subpopulation will be undertaken separately from analyses involving the Omicron Subpopulation, following the plans laid out in protocol version 7.0 and described in this SAP for the active-controlled non-inferiority Phase 3 trial. It is recognized that there will be limited power to assess non-inferiority with respect to the hospitalization/death primary outcome measure.

12.4 Data and Safety Monitoring for the Evaluation of SAB-185 under Protocol Version 8.0

In addition to the details regarding data and safety monitoring laid out in the Master Protocol, the DSMB may consider results from the “Non-Omicron Subpopulation” enrolled under protocol version 7.0 to guide their recommendations, particularly regarding any safety issues or possible early termination of the placebo-controlled evaluation of SAB-185 based on lack of sufficient

efficacy. For example, data suggesting that SAB-185 may be less effective than casirivimab plus imdevimab in the Non-Omicron Subpopulation might support a finding of lack of sufficient efficacy of SAB-185 versus placebo in the “Omicron Subpopulation”. Note, however, that a recommendation to terminate randomization in the phase III placebo-controlled superiority evaluation of SAB-185 (being conducted under protocol version 8.0) based on a finding of superiority of SAB-185 versus placebo should, in general, be based only on results from the Treated Population for this comparison defined in section 12.1 of this appendix.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, in phase II, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 72).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Phase II/III Placebo-Controlled and Phase III Active-Controlled

Study Components

Version 9.1

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 8.0 and Letter of Amendment #1

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

April 12, 2022

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**

VERSION HISTORY **5**

GLOSSARY OF TERMS **8**

1 INTRODUCTION **9**
1.1 Purpose **9**
1.2 Version History of this SAP **9**

2 STUDY OVERVIEW **11**
2.1 Study Design **11**
2.2 Randomization Process **12**
2.3 Study Objectives **13**
2.3.1 Primary Objectives 13
2.3.2 Secondary Objectives 14
2.3.3 Exploratory Objectives 14
2.4 Overview of Sample Size Considerations **15**
2.4.1 Phase II – Placebo-Controlled Superiority 16
2.4.2 Phase III – Placebo-Controlled Superiority Trial Used for the Evaluation of BRII-196+BRII-198 16
2.4.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant) 17
2.4.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 19
2.5 Overview of Formal Interim Monitoring **22**
2.5.1 Phase II – Placebo-Controlled Superiority 22
2.5.2 Phase III – Placebo-Controlled Superiority Used for the Evaluation of BRII-196+BRII-198 23
2.5.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant) 24
2.5.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 26
2.6 Graduation to Phase III **27**

3 OUTCOME MEASURES **28**
3.1 Primary Outcome Measures: Phase III **28**
3.2 Primary Outcome Measures: Phase II **28**
3.3 Secondary Outcome Measures **29**

3.4 Other Outcome Measures 32

4 STATISTICAL PRINCIPLES 34
4.1 General Considerations 34

5 ANALYSIS APPROACHES 37
5.1 Analyses of the Primary Objectives 38
5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation 38
5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation 42
5.1.2 Primary Safety (Phase II and III) 44
5.1.3 Primary Clinical Symptoms (Phase II) 45
5.1.4 Primary Virologic (Phase II) 49
5.2 Analyses of Secondary Objectives 51
5.2.1 Secondary Clinical Symptoms 52
5.2.2 Secondary Virology 57
5.3 Exploratory Analyses 60
5.3.1 New SARS-CoV-2 among Household Contacts 60
5.3.2 Hospitalization Course 60
5.3.3 Resistance Mutations 61
5.4 Interim Analysis Considerations 61

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 63

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 66
7.1 Randomization Details 66
7.2 Secondary Outcome Measures 66
7.3 Analysis Approaches 66

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 67
8.1 Secondary Outcome Measures 67
8.2 Analysis Approaches 67

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 68
9.1 Secondary Outcome Measures 68
9.2 Analysis Approaches 68

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 69
10.1 Objectives 69
10.1.1 Secondary Objectives 69
10.1.2 Exploratory Objectives 70
10.2 Outcome Measures 70
10.2.1 Secondary Outcome Measures 70
10.2.2 Other Outcome Measures 70
10.3 Analysis Approaches 70
10.3.1 Secondary Analyses 70
10.3.2 Exploratory Analyses 71

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 72
11.1 Objectives 72
11.1.1 Secondary Objectives 72
11.1.2 Exploratory Objectives 72
11.2 Outcome Measures 72
11.2.1 Secondary Outcome Measures 72
11.3 Analysis Approaches 72
12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 73
12.1 Study Design and Analysis Population Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185 73
12.2 Additional Analysis Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185 74
12.3 Analysis Considerations for the Non-Omicron Subpopulation Enrolled under Protocol Version 7.0 75
12.4 Data and Safety Monitoring for the Evaluation of SAB-185 under Protocol Version 8.0 75
13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 77
13.1 Secondary Outcome Measures 77
13.2 Analysis Approaches 77

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at 'lower' risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., 'higher' risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at 'higher risk'	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	
6.0	Update SAP to address changed implemented in protocol version 7.0 and Letter of Amendment 1. Specifically: - Updated SAP to reflect changes in study objectives, schedules for some evaluations, and associated outcome measures. - Updated SAP to reflect changes in interim monitoring, mainly related to changes in graduation criteria. - Updated this SAP to focus on the placebo-controlled phase III evaluation. (Note that a separate *revision to the* SAP will be prepared for the active-controlled phase III evaluation introduced in protocol version 7.0). [Note that the italicized text was added in version 7.0 of the SAP].	September 13, 2021
7.0	Updated SAP with the following major changes: - To add details that are specific to the active-controlled phase 3 evaluation of agents based on protocol version 7.0 and Letter of Amendment 1. - To edit some text to provide clarity concerning the analysis approaches which are the same regardless of whether a placebo control or active control is involved. In part, to achieve this, the terminology "comparator intervention" is often used. - To replace the previous section 5.4 concerning interim analysis considerations for the placebo-controlled phase III trial (which have been completed) with interim analysis considerations for the active-controlled phase III trial. - To indicate exclusion from analysis of viral shedding results from samples labelled as 'Thawed', 'Destroyed', 'Quantity Not Sufficient' or 'Invalid Specimen' as approved by the trial sponsor. - To focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.	October 24, 2021

8.0	Updated SAP to address changes implemented in Letter of Amendment 2 to protocol version 7.0. Specifically: - Added oxygen saturation as a secondary outcome in phase 3 - For the SNG001 agent, added phase II secondary objectives. - For the SNG001 agent, revised phase III exploratory objectives. - For the SNG001 agent, added details on subgroup analyses. Removed Hodges-Lehmann analysis from all analysis sections as the validity of this analysis is questionable for the type of data being generated in this study for the affected outcome measures.	January 24, 2022
9.0	Updated SAP to address changes implemented in protocol version 8.0. Specifically: - Added design details of new superiority placebo-controlled phase III trial that was introduced, including differences from the phase III superiority evaluation of the BRII agent - Outlined statistical analyses of the SAB-185 agent including implications of changing the phase III design due to the Omicron variant - Changed the Primary Symptom Duration Outcome in phase II as per change in protocol version 8.0 - Corrected the phase II “other” outcome measure for the SNG001 agent	February 25, 2022
9.1	Updated SAP to note that LOA #1 to Protocol Version 8.0 has been implemented and no changes to the Primary SAP are needed. Corrected a typo by deleting ‘cough’ from the exploratory analysis approaches section of the SNG001 appendix, as this was removed from the exploratory objective and other outcome in prior SAP versions, but was not deleted from the analyses approach section.	April 12, 2022

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as "SAP" in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III placebo-controlled investigations of ACTIV-2/A5401, as well as the phase III active-controlled investigation introduced in protocol version 7.0 and the phase III placebo-controlled investigation introduced in protocol version 8.0 . This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 stated that enrollment to all agents (except BRII-196+BRII-198 which was already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents (note that this was subsequently changed in protocol version 8.0—see below). SAP version 5.0 therefore described planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a placebo. In

addition, SAP version 5.0 addressed some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

Protocol version 7.0 introduced an open-label non-inferiority phase III design to compare investigational agents to an active-comparator among persons at higher risk of progression to hospitalization or death. This phase III evaluation was separate from the phase II superiority evaluation of agents compared to placebo among persons at lower risk for progression to hospitalization or death. Changes introduced in SAP version 6.0 focused on changes made under protocol version 7.0 (and letter of amendment #1) that related to the placebo-controlled superiority phase II/III design (note: BRII-196+BRII-198 is the only agent that enrolled in the placebo-controlled phase III design until protocol version 8.0 was introduced when a placebo-controlled design was also used for the phase III evaluation of SAB-185). Changes introduced in SAP version 7.0 addressed the introduction of the active-controlled non-inferiority phase III trial in protocol version 7.0 (and letter of amendment #1). SAP version 7.0 also introduced the exclusion from statistical analysis of results generated from problematic virologic samples based on a decision made by the DAIDS and study team. In addition, section 5.4 concerning interim analysis considerations was revised to replace considerations for the placebo-controlled phase III trial for which DSMB monitoring had been completed with considerations for the active-controlled phase III trial. Finally, adjustments were made to focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.

SAP version 8.0 implemented changes made under letter of amendment #2 to protocol version 7.0, which added oxygen saturation outcome for the active-controlled phase III and new phase III secondary and exploratory objectives for the SNG001 agent. In addition, analyses using the Hodges-Lehmann estimate were removed throughout as the validity of these analyses is questionable for the type of data being generated in this study for the affected outcome measures.

While the phase III evaluation of SAB-185 was ongoing using an active-controlled non-inferiority design, in vitro data suggested that the active control agent would not have activity against the newly emergent Omicron variant of SARS-CoV-2. As a result, enrollment to the phase III non-inferiority trial was terminated and was redesigned in order to continue a phase III evaluation of SAB-185. This design was defined in protocol version 8.0 and provided for phase III evaluation of an investigational agent versus placebo, but allowing participants to receive other COVID-19 treatments after study entry if available (availability was, however, expected to be very limited). In essence, this led to the reintroduction of a placebo-controlled phase III trial and the general approach for statistical analyses in protocol version 8.0 for this phase III trial follows the earlier plan for the placebo-controlled phase III evaluation of BRII-196+BRII-198. SAP version 9.0 was implemented to describe these changes.

Participants infected with the Omicron SARS-CoV-2 variant who were randomized under protocol version 7.0 to the “active” control agent in the phase III non-inferiority evaluation of SAB-185 were thought to have been treated with an ineffective agent, so functionally with a placebo from an efficacy perspective, The SAB-185-specific appendix of protocol version 8.0 therefore specifies that the subpopulation of participants enrolled in the non-inferiority phase III evaluation of SAB-

185 under protocol version 7.0 who were definitely or very likely infected with the Omicron variant would be included in the analysis population for the placebo-controlled evaluation of SAB-185. This particular nuance is described in more detail in the SAB-185-specific appendix of this SAP.

SAP version 9.0 also includes a change to an exploratory objective and associated outcome measure for the evaluation of SNG001 that was introduced in protocol version 7.0 but was not reflected in the applicable previous versions of the SAP.

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 8.0 for the placebo-controlled phase II and phase III evaluations of investigational agents. This section also includes a description of the non-inferiority phase III evaluation of investigational agents per protocol version 7.0 and letter of amendments 1 and 2.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient phase II and phase III testing of new agents within the same trial infrastructure.

Version 8.0 of the protocol provides for a blinded phase II evaluation of an investigational agent compared to placebo among participants at lower risk for progression to hospitalization or death, regardless of the mode of administration of the agent; for some agents, enrollment to higher risk participants in phase II was allowed under earlier protocol versions.

Based on protocol-specified criteria, agents could graduate from phase II to phase III evaluation. The phase III evaluation of investigational agents under all protocol versions has been in a population of participants at higher risk of hospitalization or death (though the definition of “higher risk” as changed across protocol versions). When protocol version 8.0 was introduced, only two agents had graduated to phase III evaluation and started enrollment: BRII-196+BRII-198 and SAB-185. For these two agents, protocol version 8.0 provides for:

- Continued follow-up of participants enrolled under protocol versions 2.0 to 6.0 into a placebo-controlled phase III trial evaluating the combination monoclonal antibody agent, BRII-196 + BRII-198.
- Continued follow-up of participants enrolled under protocol version 7.0 into a non-inferiority phase III trial evaluating the polyclonal antibody agent, SAB-185 using the monoclonal antibody combination of casirivimab plus imdevimab (REGEN-COV, Regeneron) as the control regimen. As noted above, enrollment to this non-inferiority trial was terminated because of an anticipated lack of efficacy of casirivimab plus imdevimab against the Omicron SARS-CoV-2 variant that became widely prevalent.

- Enrollment of participants into a placebo-controlled phase III trial evaluating SAB-185. In the latter trial, use of COVID-19 treatments obtained outside of the trial is allowed in both randomized arms, if available (though availability is expected to be very limited).

When two or more agents are being evaluated in the same phase of the study, the trial design includes sharing of the appropriate control group (placebo or active comparator) for efficient evaluation of each agent. Note, however, that enrollment to the phase III placebo-controlled evaluation of BRII-196+BRII-198 did not coincide with enrollment the phase III placebo-controlled evaluation of SAB-185 and so there is no sharing of the placebo control group for these two agents.

Eligible participants enrolled under all versions of the protocol from version 2.0 have intensive follow-up through day 28, followed by limited follow up through at least week 72 weeks in phase II and phase III.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

Under protocol version 8.0, the phase II trial and the phase III trial involve different populations and have separate randomizations. However, the structure of the randomization process is the same for each of the two trials, as described in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of concurrently randomized participants on a given investigational agent and in the placebo comparator group for that agent (i.e. combining participants who were eligible to receive the agent but who were randomized to any of the available placebos for investigational agents in the same phase of evaluation).

To achieve having a similar number of participants on the active arm and in the pooled comparator group for a given investigational agent, the randomization occurs in two steps within each trial.

The first randomization is to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents in the trial (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). Trial phase for an agent is accounted for in the participant eligibility (i.e. by the classification of their risk for hospitalization/death as lower or higher). In the event that a

participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

Immediately following the first randomization, participants are randomized within an Agent Group in the second randomization to the investigational agent or appropriate comparator (i.e., the matching placebo for agents in the same phase of evaluation). For a given participant, the probability of assignment to the investigational agent or placebo in the second randomization depends on the number of agents currently under investigation that the participant was eligible to receive, as phase II and phase III have distinct populations (phase II is restricted to those at lower risk of progression to hospitalization and death, and phase III is restricted to those at higher risk).

Both the first and second randomizations involve blocked stratified randomization. In phase II, both the first and second randomizations are stratified by time from symptom onset (≤5 days vs >5 days), however, in previous versions of the protocol, in which both 'higher' and 'lower' risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group ('higher' vs 'lower'). In the active-controlled phase III trial introduced in protocol version 7.0 and the placebo-controlled trial introduced in protocol version 8.0, both randomization steps are stratified by country. Under previous versions of the protocol for the placebo-controlled phase III trial evaluating BRII-196+BRII-198, both randomization steps were only stratified by time from symptom onset (≤ 5 days vs > 5 days). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 8.0; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed outside of this SAP. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed separately.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, and 14 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical

needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in NP swabs [Protocol Objective 1.2.3].

4) Phase III: To determine the efficacy of the investigational agent to increase the proportion of participants with NP SARS-CoV-2 RNA below the LLoQ at study day 3 [Protocol Objective 1.2.4]

5) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.5].

6) Phase II and III: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72 [Protocol Objective 1.2.7].

8) Phase III: To evaluate if the investigational agent reduces the time to sustained symptom resolution through study day 28 [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of hospitalization or death through study day 28, excluding hospitalizations that are determined to be unrelated to COVID-19 [Protocol Objective 1.2.9].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes and differences between investigational agent and control (placebo in phase II and active comparator in phase III) across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course in those hospitalized [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 8.0. The sample size for the placebo-controlled superiority phase III design was also the same under protocol versions 2.0 to 6.0 (it was originally defined in Appendix IV of protocol version 2.0 for the BRII-196+BRII-198 agent entered into the study under protocol version 2.0) and is currently detailed in Appendix V of protocol version 8.0 for the BRII-196+BRII-198 agent. The sample size for the non-inferiority phase III design that enrolled participants under protocol version 7.0 is detailed in section 10.4 of

protocol version 7.0. As noted above, enrollment to that non-inferiority trial was terminated early. The sample size for the placebo-controlled phase III trial introduced in protocol version 8.0 is detailed in section 10.4 of protocol version 8.0.

2.4.1 Phase II – Placebo-Controlled Superiority

For each investigational agent in phase II, the proposed sample size is 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group (range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

2.4.2 Phase III – Placebo-Controlled Superiority Trial Used for the Evaluation of BRII-196+BRII-198

The proposed sample size was 842 participants consisting of 421 participants who received the active agent and 421 participants who were concurrently randomized to placebo control. This sample size included the enrollment that occurred during the phase II placebo-controlled evaluation of an agent. Participants who were randomized but did not start their randomized investigational agent or placebo were not followed.

This sample size was chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This was based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.4.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant)

The active-controlled Phase III trial was focused on a non-inferiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus an active comparator agent, specifically the monoclonal antibody combination of casirivimab plus imdevimab. The non-inferiority margin for the absolute difference in proportion hospitalized/dead was 3% (investigational agent minus active comparator agent); the rationale for this choice was described in Section 3.1 of protocol version 7.0. Non-inferiority was considered to be established if a two-sided exact 95% confidence interval for the absolute difference was entirely below 3%. Details of the construction of the confidence interval are in section 10.6 of protocol version 7.0 and are included further below in this SAP.

The sample size differed between infused investigational agents (600 for the investigational agent and 600 for the concurrently randomized active comparator) and non-infused investigational agents (800 per arm instead of 600 per arm). The rationale for this was that there may be broader clinical utility for non-infused agents such that a slightly higher true hospitalization/death rate may be tolerated in clinical practice. No enrollment occurred for a non-infused agent and so the sample size justification described below is just for an infused agent (enrollment only occurred for the SAB-185 infused agent).

Sample Size Justification for Infused Investigational Agents

For the evaluation of a specific infused investigational agent, the sample size was 1200 participants including approximately 600 participants randomized to receive the infused investigational agent and approximately 600 participants (who were eligible to receive the infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size was chosen to provide close to 90% power to establish non-inferiority assuming that the true proportion hospitalized/dead for both the infused investigational agent and the active comparator agent was 2.3%. The rate of 2.3% was based on the observed proportion for casirivimab plus imdevimab combining across doses in the subpopulation of the Regeneron COV-2067 clinical trial who met the criteria for being at high risk of progression to hospitalization/death (FDA communication to DAIDS/NIAID, May 2021). No adjustment for loss to follow-up was made in the sample size as the primary analysis was to be based on the observed number of hospitalizations divided by the number of participants who initiated study treatment. In addition, the impact of any loss to follow-up was expected to be minimal as there was to be regular contact between research site staff and participants (or their secondary contacts) and previous experience in the study and other trials has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up).

The potential power of the study was evaluated in two ways using the PASS version 15 sample size calculation software. Both used a non-inferiority hypothesis testing approach based on use of the Miettinen and Nurminen score test statistic (which was the basis for calculating the confidence interval used for the analysis). The first ignored interim monitoring but used a binomial

enumeration method to calculate power and type I error rates. Use of the binomial enumeration method takes account of the discreteness of the binomial distribution (rather than using a normal approximation to the binomial distribution) which may be important in the setting of low hospitalization/death probabilities. Using this approach gave a power of 90.2%. The second approach did not use a binomial enumeration but took account of interim analyses using a standard implementation of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 90.0% (width of 95% confidence interval around this simulation-based value was 0.12%). Based on these two approaches, it was anticipated that the study would have had close to 90% power to show non-inferiority for an infused investigational agent assuming that it truly had the same 2.3% hospitalization/death rate as the active comparator agent.

The PASS software was also used to illustrate how the power of the study might change for various scenarios which differed from the scenario assumed (see Table below). This was undertaken using the first of the two approaches mentioned above (i.e., using the binomial enumeration approach). Looking at the top part of the table in which both the infused investigational agent and the active comparator agent have the same underlying true hospitalization/ death rate, the power is decreased if the true rate was above the assumed 2.3%, but increased if the true rate was less than 2.3%. If the true rate was 3%, then the power was still above 80%, but if the true rate is 4% it was reduced to 73%.

The middle and lower parts of the table show scenarios in which the infused investigational agent had a true hospitalization/death rate of 0.5% or 1% worse than the active comparator agent, respectively. If the true rate for the active comparator agent was 2.3% and was 2.8% for the infused investigational agent (i.e., 0.5% worse), then the power was reduced to 73%. If the true rate for the active comparator agent was 2.3% and was 3.3% for the infused investigational agent (i.e., 1% worse), then the power was reduced to 50%

Table: Power for various scenarios based on non-inferiority hypothesis testing using the likelihood score test statistic (Miettinen and Nurminen method) with binomial enumeration of power and Type I error rate. All scenarios use a 3% non-inferiority margin and one-sided Type-I error rate of 0.025 with a sample size of 600 participants receiving an infused investigational agent and 600 participants receiving the active comparator agent. Power in practice would have been slightly reduced from the values shown due to interim monitoring.

Same Underlying True Hospitalization/Death Rate For Active Comparator Agent and Infused Investigational Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
99.4%	1%	1%	2.2%
97.3%	1.5%	1.5%	2.2%
93.2%	2%	2%	2.3%
90.2%	2.3%	2.3%	2.4%
88.1%	2.5%	2.5%	2.4%
83.1%	3%	3%	2.4%
78.1%	3.5%	3.5%	2.4%
73.2%	4%	4%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 0.5% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
92.5%	1%	1.5%	2.2%
85.2%	1.5%	2%	2.2%
77.4%	2%	2.5%	2.3%
73.1%	2.3%	2.8%	2.4%
70.5%	2.5%	3%	2.4%
64.8%	3%	3.5%	2.4%
59.6%	3.5%	4%	2.4%
55.0%	4%	4.5%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 1% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
71.3%	1%	2%	2.2%
61.7%	1.5%	2.5%	2.2%
54.0%	2%	3%	2.3%
50.4%	2.3%	3.3%	2.4%
48.4%	2.5%	3.5%	2.4%
44.0%	3%	4%	2.4%
40.0%	3.5%	4.5%	2.4%
36.7%	4%	5%	2.4%
*Actual type I error rate is slightly lower than assumed rate of 2.5% because of discreteness of the binomial distribution.			

2.4.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0

The proposed sample size is 1200 participants consisting of approximately 600 participants who are randomized to receive the active agent and approximately 600 participants who are concurrently randomized to placebo control. Unlike the placebo-controlled phase III evaluation of

investigational agents under earlier versions of the protocol, under protocol version 8.0, participants enrolled in the phase II evaluation of an investigational agent are not part of the study population for the phase III evaluation of the same agent (as participants in the phase II evaluation were generally “lower risk” and participants in the phase III evaluation are “higher risk” for hospitalization/death). Participants who are randomized but do not start their randomized investigational agent or placebo are not followed.

The phase III trial is focused on a superiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus placebo (with use of SOC treatment in both arms, if available). The primary analysis will focus on evaluating the ratio of proportions (investigational agent/placebo) or, equivalently, the relative reduction in risk of hospitalization/death for the active investigational agent versus placebo. The sample size of 1200 participants, with approximately 600 randomized to an investigational agent and 600 to placebo, has been chosen to give good power (>90%) to detect relative risk reductions of 70% (as found for other antibody treatments) if the proportion hospitalized/dead in the placebo group is about 5% or higher, using a two-sided Type I error rate of 5%. There are multiple factors that will affect the power, which are discussed below. To provide context for this discussion, the table below shows the power of the study to detect relative risk reductions of between 50% and 70% for proportions hospitalized/dead in the placebo group of 3% to 6%. The powers shown were obtained in PASS software (version 15.0.4) for testing two proportions using a z-test (so the normal approximation method) with unpooled variance. They are based on an effective sample size of 570 per arm, with the 5% reduction from 600 per arm built in to allow for loss to follow-up and interim monitoring using the O’Brien and Fleming stopping guideline.

Power to detect various true effect sizes (relative reduction in risk of hospitalization/death) for selected true proportions hospitalized/dead on placebo between 3% and 6%			
Proportion Hospitalized/Dead		Relative Risk Reduction for Active versus Placebo	Power
Placebo	Active		
3%	0.9%	70%	73%
	1.2%	60%	56%
	1.5%	50%	40%
4%	1.2%	70%	85%
	1.6%	60%	69%
	2.0%	50%	51%
5%	1.5%	70%	92%
	2.0%	60%	79%

	2.5%	50%	61%
6%	1.8%	70%	96%
	2.4%	60%	86%
	3.0%	50%	69%

Discussion of Factors Affecting the Power of the Study

a. Proportion hospitalized/dead in placebo control group: As can be seen in the table, the proportion hospitalized/dead in the placebo arm has a reasonable effect on power with lower proportions leading to a reduction in power. For the placebo control group for evaluating the BRII agent in ACTIV-2, the proportion was 11% [25]. However, the proportion in the phase 3 trial of sotrovimab was 6% [26]. There is also a possibility that the proportion may be lower for the Omicron variant than with previous variants.

b. Use of SOC treatment by some participants: Higher use of SOC treatment will reduce the proportion hospitalized/dead in both randomized arms. For example, the proportion hospitalized/dead in the placebo arm would change from 6% if none receive SOC treatment to 5.58%, 4.74% and 3.90% if SOC treatment is used by a random sample of 10%, 30% and 50%, respectively, of participants in the placebo arm (i.e., SOC treatment use is not related to risk of hospitalization/death) and SOC treatment reduces risk of hospitalization/death by 70%. If SOC treatment use is not random, for example, it is taken up by the highest risk participants, then the impact might be larger. As the trial excludes participants who have accessed SOC treatment prior to entry and there is a general lack of availability of such treatments globally, use in the trial is expected to be very low (e.g., <10%) and so will limit the impact.

c. Differential effect of an investigational agent versus placebo according to use or not of SOC treatment: For a given proportion of participants hospitalized/dead in the placebo arm, the power shown in the above table is valid if the relative effect of SAB versus placebo is not affected by the use of SOC treatment. Power would be reduced from the values shown if the effect of SAB versus placebo is reduced in the presence versus absence of background therapy. A related concern arises if use of SOC treatment is differential in the investigational agent arm versus the placebo arm. For example, accessing SOC treatment at a higher rate in the placebo arm because more participants have a deteriorating health status might diminish a true difference in effect between arms and hence reduce power. As noted above, use of SOC treatment is expected to be low and so any reduction in power is expected to be limited even if this occurs.

d. Failure to start randomized treatment and loss to follow-up: The impact of any loss to follow-up is expected to be minimal as there will be regular contact between research site staff and participants (or their secondary contacts). Previous experience in the study and other trials

has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up) when loss to follow-up has also been minimal (approximately 1 to 1.5%) in this study. In addition, a very small proportion of randomized participants will not start study treatment and will be excluded from the analysis of the primary outcome. Based on ACTIV-2 experience, allowance for 3-4% not starting treatment or being lost to follow-up before hospitalization is built into the above power table (with additional allowance of 1-2% for interim monitoring using the O'Brien and Fleming stopping guideline).

Because of these uncertainties, the DSMB will be asked to monitor the potential impact of the above factors on the operational feasibility of the study.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring of the placebo-controlled superiority phase II, the placebo-controlled superiority phase III for evaluating BRII-196+BRII-198, the active-controlled non-inferiority phase III for evaluating SAB-185 under protocol version 7.0, and the placebo-controlled superiority phase III for monitoring investigational agents under protocol version 8.0. Additional details on phase II monitoring can be found in protocol version 8.0 section 10.5.1, and in protocol version 8.0 Appendix V for placebo-controlled phase III monitoring for the BRII-196+BRII-198 agent. Details on active-controlled non-inferiority phase III monitoring for SAB-185 are taken from protocol version 7.0 section 10.5.2 as amended in letter of amendment 1 to protocol version 7.0. Details on the placebo-controlled superiority phase III monitoring introduced under protocol version 8.0 are taken from section 10.5.2 of protocol version 8.0. Statistical considerations for interim monitoring are described in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to study product or if two participants experience a Grade 4 AE deemed related to study product, enrollment to the study product group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II – Placebo-Controlled Superiority

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III – Placebo-Controlled Superiority Used for the Evaluation of BRII-196+BRII-198

[At the time of preparing this version of the SAP, all participants randomized to receive BRII-196+BRII-198 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study].

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

2.5.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant)

[At the time of preparing this version of the SAP, all participants randomized in the non-inferiority phase III evaluation of SAB-185 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study. Note that the first interim analysis for this part of the study was scheduled before the termination of enrollment due to the anticipated lack of efficacy on the control regimen against the SARS-CoV-2 Omicron variant that had become widely prevalent. That interim analysis was replaced by an interim analysis that followed the monitoring approach described in Section 2.5.4 and the SAB-185-specific appendix of this SAP].

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if there is persuasive evidence of non-inferiority (or superiority or inferiority) of an investigational agent versus the active comparator agent in its effect on the hospitalization/death outcome. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend

termination or modification of the study if it appears futile on statistical or operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, three interim analyses for DSMB review are planned for each investigational agent, after approximately 25%, 50% and 75% of the planned enrollment for an investigational agent has been completed and followed through to day 28. At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guidelines for Efficacy or Lack of Efficacy

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses) for the absolute difference (investigational agent minus active comparator agent) in the proportion of participants hospitalized or dead by day 28, relative to thresholds defining non-inferiority, superiority or inferiority of the investigational agent as follows (in the order given):

- The DSMB may recommend releasing results evaluating the effect of an investigational agent when both non-inferiority and superiority of that agent is established based on the confidence interval being entirely below 0% (i.e., supportive of a lower true proportion being hospitalized or dying on the investigational agent than the active comparator agent). If this occurs, consideration will need to be given to the ongoing appropriateness of the active comparator agent as a control for evaluating other investigational agents in the study.

- Early stopping and/or release of results based on non-inferiority should be considered on an agent-by-agent basis. For non-infused agents, the DSMB may recommend releasing results evaluating the effect of an investigational agent when non-inferiority (but not superiority) of that agent is established based on the confidence interval being entirely below 3% (but not entirely below 0%). However, in the interests of also having an adequate safety database for the investigational agent, it is not intended that this recommendation be made before approximately 400 participants have been randomized to receive the agent (or some other number of participants specified in the agent-specific appendix). In addition, the study may continue randomizing participants to the investigational agent in the interests of increasing precision in evaluating the agent; this decision will be made by the study team and sponsor on an agent-by-agent basis. For infused agents, early stopping and/or release of results for non-inferiority should not be considered.

- The DSMB may recommend releasing results and terminating randomization to an investigational agent if inferiority of that agent is established based on the confidence interval being entirely above 0% (i.e., suggesting a higher true proportion being hospitalized or dying on the investigational agent than the active comparator agent). Examples of how this criterion might be met when evaluating an infused agent and when

the observed control rate is close to 2.3% include observing 18/150 versus 3/150 (observed difference 10.0%) at the first interim analysis; 23/300 versus 7/300 (observed difference 5.3%) at the second interim analysis; and 24/450 versus 10/450 at the third interim analysis (observed difference 3.1%). In these examples, all observed differences are higher than the non-inferiority margin of 3%, and are indicative also of the futility of continuing evaluation of the infused investigational agent to demonstrate non-inferiority.

2.5.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if efficacy of the agent versus placebo has been established, or if it is unlikely that the agent has sufficient efficacy to warrant further evaluation in this study. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend termination or modification of the study if it appears futile on operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, two interim analyses for DSMB review are planned for each investigational agent, after approximately one-third (i.e., approximately 400 participants) and two-thirds (i.e., approximately 800 participants) of the planned enrollment for an investigational agent has been completed and followed through to day 14 (the choice of day 14 is because the large majority of hospitalizations/deaths in ACTIV-2 have been observed to occur by day 14).

At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guideline for Efficacy Favoring an Investigational Agent versus Placebo

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses—see further below) for the relative difference (investigational agent / placebo) in the proportion of participants hospitalized or dead by day 28. As a stopping guideline for greater efficacy of an investigational agent compared with placebo, the O'Brien and Fleming boundary will be used. The stopping guideline will be implemented using the Lan-DeMets spending function approach to allow for the possibility of changes in the timing of interim analyses and/or additional (or fewer) interim analyses if recommended by the DSMB. Information time for the spending function will be based on the proportion of the planned enrollment (i.e., of the 1200 participants for comparing an investigational agent to placebo) who could have been followed through day 14 at the time of the data freeze for the interim analysis. The choice of day 14 here reflects the fact that the very large majority of hospitalizations and deaths in ACTIV-2 have occurred by 14 days of follow-up. As a guideline, if the two-sided 95% confidence interval (adjusted for interim analyses) excludes a risk

ratio of one (equivalently a relative risk reduction of zero) favoring the investigational agent, then the DSMB may recommend closure of randomization to that agent; release of interim results may also be recommended.

There is the possibility that differences between the treatment groups may be observed early in follow-up. However, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the treatment group comparisons will be on the cumulative proportion hospitalized/dead at day 28.

Stopping Randomization to an Investigational Agent Because of Limited Efficacy

Because there are treatments available that may substantially reduce the risk of hospitalization/death, albeit with limited availability and the caveat that they have generally been evaluated among individuals infected with earlier variants of SARS-CoV-2, it is likely that a treatment which reduces the risk of hospitalization/death by less than 30% versus placebo will have limited utility in clinical practice. Therefore, as a non-binding guideline, the DSMB may recommend early termination of randomization to a specific investigational agent because of limited efficacy if the two-sided 95% confidence interval (adjusted for interim analyses) for the risk ratio is entirely above 0.7 or, equivalently, the two-sided 95% confidence interval (adjusted for interim analyses) for the relative risk reduction is entirely below 30%.

Modifying or Stopping the Study for Operational Futility

The DSMB will also monitor operational futility, in particular related to losses to follow up, low hospitalization/death rate in the placebo arm (which, in part, may arise due to more extensive use of SOC treatment than anticipated). As most hospitalizations are expected to occur early in follow up (e.g., during the first 14 days), early losses to follow up would be most relevant. As a benchmark, an overall loss to follow-up rate (excluding losses after a participant is hospitalized) of more than 5% would be cause for concern.

With regard to the hospitalization/death rate in the placebo arm, the power of the study is limited if this rate is below 3% (see power analysis table above). Therefore, as a benchmark, an observed rate of less than 3% in the placebo arm would be a cause for concern. If this arises, or temporal trends in hospitalization/death rate suggest it might, then any DSMB recommendation concerning this issue might incorporate information about factors that might be driving it (e.g., increasing use of SOC treatment, evolving lower risk of participants enrolled, or lower risk with new variants).

2.6 Graduation to Phase III

[At the time of preparing this version of the SAP, all agents that had been initiated in phase II evaluation have been evaluated for graduation to phase III evaluation or a decision has been made not to evaluate them for graduation. The following therefore provides a summary of the approach to graduation that is in protocol version 8.0 recognizing that there are currently no further agents being considered for graduation].

Each investigational agent that is being considered for evaluation in phase III will be evaluated for safety, for activity in reducing COVID-19 symptoms and hospitalization/death, and for activity in reducing SARS-CoV-2 RNA shedding. An analysis to determine if an agent should graduate from phase II, and enter phase III, will be conducted when 220 participants assigned to the agent or concurrent placebo in phase II evaluation have completed their Day 7 evaluations and have the required data available in the database. Additional interim graduations may be assessed for some agents, see agent-specific appendices in the protocol for details.

The DSMB will review unblinded data and make recommendations to NIAID (as trial sponsor) and to the TOC, indicating whether graduation criteria have been met. The recommendation for an agent to enter phase III evaluation will be made by the TOC in discussion with the collaborating company; the collaborating company that is responsible for the agent will decide whether or not to adopt the recommendation. The TOC and collaborating company will also consider which dose to recommend for evaluation in phase III, for investigational agents with more than one dose under evaluation in phase II. NIAID/DAIDS, as the sponsor of the study, will make the final determination regarding graduation of the study product.

The TOC may recommend an agent move directly into phase III, without evaluation in phase II in ACTIV-2, if there is sufficient safety and efficacy data supporting phase III evaluation available from outside of the trial. These agents will not undergo graduation analyses.

Graduation criteria and statistical considerations are discussed in the Graduation Rules SAP.

3 Outcome Measures

All outcome measures are copied from the protocol version 8.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo [active comparator intervention under protocol version 7.0]. [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the number of days from start of investigational treatment to the first of two consecutive days when all symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: At each of days 3, 7 and 14 quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs. [For Primary Objective 3]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
   [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
   [Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at

baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment. [Supportive of Primary Objective 2]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment. [Supportive of Primary Objective 2 and for Secondary Objective 8]

   Duration defined as the number of days from start of investigational treatment to the first of four consecutive days when all symptoms are scored as absent. Targeted symptoms as defined in the primary phase II clinical (symptom duration) outcome.

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28. [Supportive of Primary Objective 2]

   Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of two consecutive days that a participant reported return to usual (pre-COVID) health.

7) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28. [Supportive of Primary Objective 2]

   Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

8) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention, hospitalization, and death. [For Secondary Objective 1].

   Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

9) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or comparator intervention. [For Secondary Objective 2]

10) Phase II and Phase III: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

11) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

Virology

12) Phase III (Active-Controlled [protocol version 7.0] and placebo-controlled [protocol version 8.0]) only: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs at day 3. [Support of Primary Objective 3]

13) Phase II and III: Level (quantitative) of SARS-CoV-2 RNA from staff-collected NP swabs at days 3, 7, and 14 in phase II and at day 3 in phase III.8.
[For Secondary Objective 3]

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from staff-collected NP swabs at days 0, 3, 7, and 14. [Supportive of both Primary Objective 3 and Secondary Objective 3]

Efficacy

15) Phase II only: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

17) Phase II and III: Death due to any cause or hospitalization due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death due to any cause or hospitalization due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

19) Phase II and III: Death due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

20) Phase II and III: Death due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

21) Phase III: Death due to any cause or hospitalization due to any cause, excluding hospitalizations that are deemed unrelated to COVID-19, during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 9]

    Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic. An Adjudication committee is evaluating the relatedness of hospitalization due to COVID-19.

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or comparator intervention. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

    Ordinal scale defined as:

    Death
    Hospitalized, on invasive mechanical ventilation or ECMO;
    Hospitalized, on non-invasive ventilation or high flow oxygen devices;
    Hospitalized, requiring supplemental oxygen;
    Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to its comparator intervention (placebo or active comparator, as appropriate for the agent and phase of evaluation).

- Treated Population: All participants in the Randomized Population who received any investigational agent or its comparator agent (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus its comparator intervention. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to the comparator intervention, who were also eligible to have received the investigational agent of interest. For the placebo-controlled trials, the comparison group will pool across all relevant placebos (i.e. including the placebo for the agent of interest and the placebos for other agents). For the primary placebo-controlled analysis of a specific investigational agent, a supplemental analysis may be undertaken that restricts the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/comparator intervention occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/comparator intervention.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24).

Week 72: Key visit for evaluation longer-term efficacy and safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLoQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

Virology results generated from specimens with the following conditions reported in the database will be excluded from analyses:

- Thawed;
- Invalid Specimen;
- Quantity Not Sufficient;
- Destroyed.

Note: Samples with the condition code ‘NOT’ were also to be excluded per the trial sponsor but this code indicates that the specimen was not tested. Thus, no result is expected and no exclusion is needed.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or comparator intervention are not followed.

Participants who have protocol violations, such as those who start investigational agent or comparator intervention outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at “higher” risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the placebo-controlled superiority design. Further details are provided after the table.

<table>
<tr><td colspan="2">Phase III Primary Objective for Efficacy—Placebo-Controlled Superiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.</td></tr>
<tr><td>Estimand description</td><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td>Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td>Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td>Treated Population</td></tr>
<tr><td>Variable(s)</td><td>Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td>Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td>Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td>Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td>Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td>Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="2">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5), and to 7 days under protocol version 6 and subsequent protocol versions.</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (due to any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17, 18, 19, 20 and 21, evaluate the proportion of participants who are hospitalized or died through week 24, the proportion who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days) or through 504 days (i.e. 72 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher’s exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the active-controlled non-inferiority design. Further details are provided after the table.

Phase III Primary Objective for Efficacy—Active-Controlled Non-Inferiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.	
Estimand description	Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.
Treatment	Investigational agent or active comparator agent (casirivimab and imdevimab).

Target population	Analysis set (analysis population)
Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry	Treated Population
Variable(s)	**Outcome measure(s)**
Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent (coded as 1 if participant died or was hospitalized, and 0 otherwise).	Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent.
Handling of intercurrent events	**Handling of missing data**
None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of the agent to which they were randomized.	Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as not having an event after the date last known to be alive.
Population-level summary measure	**Analysis approach**
Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization over 28 days.	Difference (for investigational agent minus active comparator agent) of the proportion dying or being hospitalized at day 28. See text for further details.

Analysis Approach

The analysis of the primary efficacy outcome in phase III will evaluate the absolute difference in proportion of participants hospitalized (due to any cause) or died (due to any cause), from day 0 through day 28, between randomized arms; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included.

Inference will be based on constructing a two-sided exact 95% confidence interval for the absolute difference in proportions (proportion for the investigational agent minus the proportion for the active comparator agent). If this confidence interval is entirely below the non-inferiority margin of 3%, then a conclusion of non-inferiority of the investigational agent compared with the active

comparator agent will provide reasonable evidence that the investigational agent is effective against COVID-19.

The exact 95% confidence interval will be calculated using the method of Chan and Zhang [Biometrics 1999;55:1201-09] as implemented, for example, in StatXact PROC BINOMIAL for SAS [StatXact 12 PROCs for SAS Users Manual. Cytel Inc., Cambridge, MA; 2019]. This method inverts two one-sided hypothesis tests (with one-sided error rate of 0.025 each) to obtain the confidence interval so providing a confidence interval-based method which preserves the type I error rate in establishing non-inferiority to be 0.025. To preserve confidence interval coverage (and type I error rate for assessing non-inferiority) over multiple interim analyses, the confidence interval will be calculated using a "repeated" confidence interval approach with spending of error rate at each interim analysis using the Land and DeMets approach with an O'Brien and Fleming spending function.

In essence, basing the comparison of treatment groups on the simple proportion of participants who were hospitalized or died assumes that participants who are lost to follow-up before 28 days without prior hospitalization were not hospitalized and did not die by 28 days. The decision to use the simple proportion for analysis rather than use, for example, a Kaplan-Meier estimate of the cumulative proportion of participants hospitalized or dying during the first 28 days of follow-up to account for losses to follow-up was taken for multiple reasons. First, in ACTIV-2 and other COVID-19 trials, most hospitalizations and deaths occur during the first two weeks of follow-up and the study has been designed to have regular contact with participants or their secondary contacts so as to maximize ascertainment of hospitalization and death information. Second, loss to follow-up has been low in the ACTIV-2 study: approximately 3% among higher risk participants. Third, with the very low rates of hospitalization/death expected (e.g., 2.3% for the active comparator agent), confidence interval coverage (and type I error rates) are better preserved at their desired levels through the use of exact statistical methods for analyzing proportions than is achieved using asymptotic statistical methods based on Wald-type analyses using Greenwood's formula to obtain standard errors for Kaplan-Meier estimates. To assess the potential impact of loss to follow-up (assumed to be non-informative) on the interpretation of results, the following sensitivity analyses will be undertaken, repeating the primary analysis repeated with:

(a) a comparison of the simple proportions using a Wald-based confidence interval; and

(b) a comparison of proportions estimated using Kaplan-Meier methods (with censoring of follow-up at the earlier of day 28 and the time that a participant was last known to be alive) using a Wald-based confidence interval with standard error based on Greenwood's formula.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 16, 17, 18, 19, 20 and 21 evaluate the proportion of participants who die through to day 28, the proportion who are hospitalized or died through week 24, the proportion

who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In the sensitivity analyses based on Kaplan-Meier estimates, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days for outcomes through to 24 weeks) or through 504 days (i.e. 72 times 7 days for outcomes through to 72 weeks).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approach outlined for the primary analysis will be implemented for each subgroup. However, these analyses are likely to involve small numbers of events in most or all subgroups and hence have very limited precision. Because of this, any assessment of treatment by subgroup interaction, if undertaken, will be considered exploratory. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/active comparator Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or comparator intervention) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

In placebo-controlled evaluations, because some agents may be administered using injections or infusions and others will not be, the primary safety analysis may be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed in the same manner as the primary safety outcomes.

Additional longer-term safety outcomes may be assessed, see agent-specific appendices for details.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the number of days from start of investigational agent (day 0, pre-treatment) to the first of two consecutive days when all symptoms scored as moderate

or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and comparator intervention arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic

missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to “absent”, and (b) four consecutive days with resolution of all targeted symptoms to “absent” (i.e., secondary outcome measure 5). For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are “absent” on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis maybe repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to comparator intervention to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) ‘Risk of Severe Disease’ Stratification [this may not be pursued for agents which predominantly enrolled participants who were at ‘lower’ risk for severe COVID progression]
5) Age Group (<60, ≥60)

6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, and 14).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as "0" if above LLoQ, and "1" if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or comparator intervention) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the number of days from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;
3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model

may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

Duration of Symptoms

In support of the symptom duration outcome in phase III, the analysis will be repeated using the same approach described in the primary symptom duration analysis for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to “absent.” To address secondary objective 8, and in supportive of the symptom duration outcome in phase III, a similar TTE outcome measure will also be examined defined as time to four consecutive days with resolution of all targeted symptoms to “absent,” (i.e. secondary outcome measure 5).

Return to Usual Health

The analysis of return to usual health will be repeated using the same approach described above for a similar TTE outcome measures defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2.2 Secondary Virology

The schedule of evaluations in protocol version 7.0 indicates that only NP swabs will collected in both phase II and phase III, and therefore only analyses of SARS-CoV-2 RNA from NP swabs are outlined below. Some agents may have completed enrollment in phase II prior to implementing protocol version 7.0, and therefore may have additional specimens collected for SARS-CoV-2 RNA testing. If analyses of these additional specimens are pursued, then the approach will be as defined in the relevant previous version of the SAP.

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA at day 3 (this is a secondary outcome for the active-controlled phase 3 only)

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from staff-collected NP swabs at entry and day 3.

The proportion of participants with SARS-CoV-2 RNA < LLoQ day 3 will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. The model will include a main effect for treatment and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained by taking the exponential of the treatment parameter estimate (and associated 95% CI). In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ). Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time for staff-collected NP swabs.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

In phase II only, levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs will be analyzed using participant-specific AUCs. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and

above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: max[0, $\log_{10}$(RNA)-$\log_{10}$(LLoQ)], obtained at the scheduled measurement times between and including day 0 and day 14.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 14 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 14 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent comparator intervention through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent comparator intervention through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent comparator intervention who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with

frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 72.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

Interim analyses of the placebo-controlled superiority phase III evaluation of an agent was finished at the time of finalization of SAP version 7.0. The following from protocol version 7.0 describes the interim analysis considerations for the active-controlled non-inferiority phase III evaluation of an agent.

The two-sided 95% confidence interval mentioned above [see section 2.5.3 of the SAP] will be adjusted for the multiple interim analyses to preserve the confidence interval coverage to at least 95% (this is also referred to as using “repeated” confidence intervals).

The standard Lan and DeMets approach will be used to achieve this, incorporating an O’Brien and Fleming spending function. For simplicity, the information scale for the spending function will be determined as the proportion of the planned enrollment randomized to the investigational agent being evaluated at the time of the interim analysis. As an example, if in practice, the analyses were after exactly 25%, 50%, 75% and 100% of the planned enrollment, then the nominal confidence intervals used to assess efficacy would have coverage 99.9985% at the first analysis, 99.70% at the second analysis, 98.17% at the third analysis and 95.60% at the fourth analysis (these were obtained from PASS software). However, as the O’Brien and Fleming spending function is very conservative at early interim analyses, making stopping very difficult, for the assessment of inferiority of an investigational agent compared to the active comparator agent, an asymmetric approach will be used to reduce the level of evidence required for early stopping in the event that an investigational agent appears inferior to the active comparator agent. Specifically, if a nominal confidence interval with coverage of greater 99.9% at an early interim analysis is suggested by use of the O’Brien and Fleming spending function, then a nominal confidence interval with coverage of 99.9% will be used instead for assessing inferiority of the investigational agent.

The DSMB will also monitor the proportion hospitalized/dead in the active comparator arm as this key parameter, coupled with the non-inferiority margin, underpins the study design. The study is designed assuming that the underlying true proportion of participants on the active comparator agent is 2.3%. This is the proportion (32/1392) observed for high risk participants in the Regeneron COV-2067 trial for the agent (pooling across doses studied in that trial; FDA communication to DAIDS/NIAID). A 95% confidence interval for this proportion is (1.5%, 3.1%).

An assessment of non-inferiority in this study would be more difficult if the proportion of participants on the active comparator agent in this study is somewhat different from that in the Regeneron COV-2067 (e.g., somewhat outside of the range suggested by the confidence interval).

For example, this might arise if variants of SARS-CoV-2 are present in the study population which the active comparator agent is less effective against. Such an issue would undermine the use of a 3% non-inferiority margin in this study. It may however be addressed by focusing the non-inferiority assessment on the subpopulation in this study without such variants (assuming these have been identified), or in establishing superiority of the investigational agent in the overall study population. This may require a larger sample size to maintain power.

Another potential reason for a somewhat different proportion hospitalized/dead on the active comparator agent in this study versus that in the Regeneron COV-2067 study is that this study is likely to enroll in a number of different countries, whereas the Regeneron COV-2067 enrolled primarily in the United States. Aside from possible differences in circulating variants among countries, differences among countries in clinical practice and/or in the availability of hospital care might lead to differences in hospitalization/death rates. The DSMB will monitor descriptive results by country and provide guidance about countries with notably low or high rates of hospitalization/death.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during

hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued through to protocol version 6.0. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions. All participants enrolled to evaluate BRII-198+BRII-196 were randomized to active agent or placebo and so placebo is mentioned as the comparator intervention throughout this appendix.

7.1 Randomization Details

Phase III is stratified by time from symptom onset (≤ 5 days versus > 5 days).

7.2 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.3 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 72).

8 Appendix 3: Statistical Considerations for AZD7442 IV

NOTE: AZD7442 IV is only being evaluated in this study in phase II with a placebo control.

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 72).

9 Appendix 4: Statistical Considerations for AZD7742 IM

NOTE: AZD7442 IM is only being evaluated in this study in phase II with a placebo control.

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 72).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

2) Phase III: To determine whether SNG001 reduces hospitalization or death through study day 28 among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

3) Phase III: To determine whether SNG001 reduces duration of targeted COVID-19-associated symptoms through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

4) Phase III: To determine whether SNG001 reduces COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

5) Phase III: To determine whether SNG001 reduces progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at entry, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

6) Phase III: To determine whether SNG001 increases proportion of individuals with pulse oximetry measurement of ≥ 96% through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

7) Phase III: To determine whether SNG001 reduces the time to sustained symptom resolution through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

8) Phase III: To determine whether SNG001 prevents the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72, among individuals in the subgroup who report moderate or severe shortness of breath or

difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

9) Phase III: To determine whether SNG001 prevents the composite endpoint of hospitalization or death through stay day 28, excluding hospitalizations that are determined to be unrelated to COVID-19, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

10.1.2 Exploratory Objectives

1) Phase II and III: To determine whether SNG001 reduces severity of shortness of breath or difficulty breathing through study day 28.

2) Phase II: To determine whether SNG001 reduces a COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 among individuals who report moderate to severe shortness of breath or difficulty breathing at day 0.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II: Area under the curve of *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *shortness of breath or difficulty breathing* symptom in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

10.3.1 Secondary Analyses

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in

a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

The secondary objectives addressing analyses among people who reported moderate or severe shortness of breath or difficulty breathing at day 0, and among people who reported severe shortness of breath or difficulty breathing at day 0, will be undertaken in the same manner as the analyses of this outcomes among the overall study population.

10.3.2 Exploratory Analyses

Exploratory analyses will compare the AUC for shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate.

To address SNG001-specific exploratory objective 2, the COVID-19 severity ranking outcome will compared between arms using the same methods outlined for the secondary analysis of this outcome measure, but restricted to be among those with moderate to severe shortness of breath or difficult breathing at day 0,

11 Appendix 6: Statistical Considerations for Camostat

Note: camostat is only being evaluated in this study in phase II with a placebo control.

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration in phase II (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a separate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

However, there are key special study design and analysis considerations for the phase III evaluation of SAB-185 (3,840 Units/kg). These are described in this appendix based on material presented in sections 3.4 and 10 of protocol version 8.0.

12.1 Study Design and Analysis Population Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185

Phase III evaluation of SAB-185 was initiated under protocol version 7.0 in a non-inferiority comparison of SAB-185 to an active control of casirivimab plus imdevimab. While enrollment was ongoing, the Omicron variant of SAR-CoV-2 became highly prevalent. In vitro data suggested that casirivimab plus imdevimab would be ineffective against this variant, and FDA authorization for emergency use of this regimen for treatment of COVID-19 was withdrawn due to non-susceptible SARS-CoV-2 variants (such as Omicron). Because of this, enrollment into the study was paused pending development of protocol version 8.0, which replaces the non-inferiority evaluation of investigational agents compared to casirivimab plus imdevimab with a placebo-controlled superiority design allowing for the additional use of standard of care (SOC) treatments, if available, in both arms.

Over 700 participants were enrolled under protocol version 7.0 and randomized to SAB-185 or casirivimab plus imdevimab. These participants can be divided into two subpopulations: (1) participants who were definitely or very likely infected with the Omicron SARS-CoV-2 variant (“Omicron Subpopulation”); and (2) participants who were definitely not, or likely not infected with the Omicron variant (“Non-Omicron Subpopulation”). Following the details in section 10.1.1 of protocol version 8.0, these two subpopulations are defined in more detail:

- The “Omicron Subpopulation” enrolled under protocol version 7.0 is defined as (1) all participants randomized under protocol version 7.0 who were infected with the Omicron variant as identified on sequencing of an NP sample obtained on day 0, plus (2) all participants randomized under protocol version 7.0 on or after December 26, 2021, who do not have variant information available from a sample obtained on day 0. The second of these two groups of participants are assumed very likely to be infected with the Omicron variant on the basis that prevalence of the Omicron variant in the U.S. was estimated by the CDC to be 89.2% for the week ending January 1, 2022 (and starting December 26, 2021), 95.2% for the week ending January 8, 2022, 97.9% for the week ending January 15, 2022, and 98.9% for the week ending January 22, 2022, during which enrollment under protocol version 7.0 was stopped [COVID Data Tracker. United States Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#variant-proportions. Accessed February 11, 2022].

- The “Non-Omicron Subpopulation” enrolled under protocol version 7.0 is defined as all participants enrolled under protocol version 7.0 excluding those in the “Omicron Subpopulation.” It therefore includes all participants randomized on or before December 25, 2021, who did not have the Omicron variant identified in sequencing of an NP sample obtained on day 0 (including those with no sequence available from an NP sample obtained on day 0). It also includes all participants randomized on or after December 26, 2021 who had a non-Omicron variant identified in sequencing of an NP sample obtained on day 0.

Based on in vitro data, the combination of casirivimab plus imdevimab is thought to have no effect on hospitalization/death in the Omicron Subpopulation and so is considered functionally to be a placebo from an efficacy perspective. Therefore, for the purposes of evaluating the superiority of SAB-185 versus placebo under this version of the protocol, the Randomized Population will be comprised of the Omicron Subpopulation enrolled under protocol version 7.0 as well as the population enrolled under protocol version 8.0 that is randomized to SAB-185 or its appropriate placebo control group. The planned sample size for this Randomized Population combining participants in the Omicron Subpopulation enrolled under protocol version 7.0 and participants enrolled under this version of the protocol is 1200 participants. The Treated Population (modified intent-to treat [mITT] population) for evaluating SAB-185 is this Randomized Population after excluding any participants who did not receive study product (i.e. excluding those who did not start SAB-185 or casirivimab plus imdevimab if enrolled under protocol version 7.0, and excluding those who did not start SAB-185 or placebo if enrolled under protocol version 8.0).

The Non-Omicron Subpopulation enrolled under protocol version 7.0 is not part of the phase III placebo-controlled superiority evaluation of SAB-185. See section 12.4 of this appendix for analysis considerations for this population.

12.2 Additional Analysis Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185

Data from the Treated Population defined in section 12.1 of this appendix will be used to evaluate the efficacy and safety of SAB-185 versus placebo (with possible SOC treatment added in both arms, if available). These analyses will follow the analysis plans in this SAP for the phase III placebo-controlled superiority evaluation of an investigational agent.

For the primary efficacy outcome measure (hospitalization/death), results comparing randomized arms may be presented separately for the Omicron Subpopulation enrolled under protocol version 7.0 and the population enrolled under protocol version 8.0. The possibility of heterogeneity in the effect of SAB-185 versus casirivimab plus imdevimab in the Omicron Subpopulation (considered functionally to be a placebo in this population) versus the effect of SAB-185 versus placebo (with the possibility of SOC treatment, if available) among participants enrolled under protocol version 8.0 may also be evaluated to assess the possible impact on interpretation of results. This analysis will follow the subgroup analysis plan for the primary efficacy outcome described in section 5.1.2 of this SAP.

For the major secondary outcome measures (4) symptom duration, (12) SARS-CoV-2 <LLoQ versus ≥LLoQ, and (13) quantitative SARS-CoV-2 RNA levels, results comparing randomized arms may also be presented separately for the Omicron Subpopulation enrolled under protocol version 7.0 and the population enrolled under protocol version 8.0. These analyses will follow the subgroup analysis plan for these secondary outcome measures described in section 5.2 of this SAP.

In addition, safety analyses will be presented separately for the following mutually exclusive subgroups:

(1) The Omicron Subpopulation enrolled under protocol version 7.0, as the control group received casirivimab plus imdevimab;

(2) Participants enrolled under protocol version 8.0, as the control group received placebo.

An additional breakdown of subgroup (2) will be undertaken for safety analyses as some participants may have received SOC treatment in addition to randomized SAB-185 or placebo:

(2a) Participants enrolled under protocol version 8.0 who did not receive SOC treatment; and

(2b) Participants enrolled under protocol version 8.0 who received SOC treatment.

It is recognized that the comparisons in subgroups (2a) and (2b) may not be pure randomized comparisons because receipt of SOC treatment may be influenced by the clinical status of a participant after randomization.

12.3 Analysis Considerations for the Non-Omicron Subpopulation Enrolled under Protocol Version 7.0

Follow-up of the Non-Omicron Subpopulation enrolled under protocol version 7.0 will continue per protocol. In this subpopulation, the combination of casirivimab plus imdevimab is expected to be effective. Analysis of outcomes from the Non-Omicron Subpopulation will be undertaken separately from analyses involving the Omicron Subpopulation, following the plans laid out in protocol version 7.0 and described in this SAP for the active-controlled non-inferiority Phase 3 trial. It is recognized that there will be limited power to assess non-inferiority with respect to the hospitalization/death primary outcome measure.

12.4 Data and Safety Monitoring for the Evaluation of SAB-185 under Protocol Version 8.0

In addition to the details regarding data and safety monitoring laid out in the Master Protocol, the DSMB may consider results from the “Non-Omicron Subpopulation” enrolled under protocol version 7.0 to guide their recommendations, particularly regarding any safety issues or possible early termination of the placebo-controlled evaluation of SAB-185 based on lack of sufficient

efficacy. For example, data suggesting that SAB-185 may be less effective than casirivimab plus imdevimab in the Non-Omicron Subpopulation might support a finding of lack of sufficient efficacy of SAB-185 versus placebo in the “Omicron Subpopulation”. Note, however, that a recommendation to terminate randomization in the phase III placebo-controlled superiority evaluation of SAB-185 (being conducted under protocol version 8.0) based on a finding of superiority of SAB-185 versus placebo should, in general, be based only on results from the Treated Population for this comparison defined in section 12.1 of this appendix.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, in phase II, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 72).

Approvals

ACTIV-2/ACTG A5401

Primary Statistical Analysis Plan

Phase II/III Placebo-Controlled and Phase III Active-Controlled

Study Components

Version 10.0

Adaptive Platform Treatment Trial for Outpatients with COVID-19

(Adapt Out COVID)

Based on Protocol Version 8.0 and Letter of Amendment #1

(Applies to Agents Entered into ACTIV-2 in Protocol Versions 2.0, 3.0, 4.0 & 5.0)

ClinicalTrials.gov Identifier: NCT04518410

January 9, 2023

Created by:

<Authors Redacted>

Harvard T.H. Chan School of Public Health

Table of Contents

TABLE OF CONTENTS **2**
VERSION HISTORY **5**
GLOSSARY OF TERMS **9**
1 INTRODUCTION **10**
1.1 Purpose **10**
1.2 Version History of this SAP **10**
2 STUDY OVERVIEW **12**
2.1 Study Design **12**
2.2 Randomization Process **14**
2.3 Study Objectives **15**
2.3.1 Primary Objectives 15
2.3.2 Secondary Objectives 15
2.3.3 Exploratory Objectives 16
2.4 Overview of Sample Size Considerations **17**
2.4.1 Phase II – Placebo-Controlled Superiority 17
2.4.2 Phase III – Placebo-Controlled Superiority Trial Used for the Evaluation of BRII-196+BRII-198 18
2.4.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant) 18
2.4.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 (Terminated Early Based on the Recommendation of the DSMB Due to Operational Futility) 22
2.5 Overview of Formal Interim Monitoring **24**
2.5.1 Phase II – Placebo-Controlled Superiority 25
2.5.2 Phase III – Placebo-Controlled Superiority Used for the Evaluation of BRII-196+BRII-198 25
2.5.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant) 27
2.5.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 (Terminated Early Based on the Recommendation of the DSMB Due to Operational Futility) 28
2.6 Graduation to Phase III **30**
3 OUTCOME MEASURES **31**
3.1 Primary Outcome Measures: Phase III **31**

3.2 Primary Outcome Measures: Phase II 31
3.3 Secondary Outcome Measures 32
3.4 Other Outcome Measures 35

4 STATISTICAL PRINCIPLES 37
4.1 General Considerations 37

5 ANALYSIS APPROACHES 40
5.1 Analyses of the Primary Objectives 41
5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation 41
5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation 45
5.1.2 Primary Safety (Phase II and III) 47
5.1.3 Primary Clinical Symptoms (Phase II) 48
5.1.4 Primary Virologic (Phase II) 52
5.2 Analyses of Secondary Objectives 54
5.2.1 Secondary Clinical Symptoms 55
5.2.2 Secondary Virology 60
5.3 Exploratory Analyses 63
5.3.1 New SARS-CoV-2 among Household Contacts 63
5.3.2 Hospitalization Course 63
5.3.3 Resistance Mutations 64
5.4 Interim Analysis Considerations 64

6 APPENDIX 1: ALGORITHM FOR HANDLING MISSING SYMPTOM EVALUATIONS FOR THE PRIMARY PHASE II SYMPTOM OUTCOME MEASURE 66

7 APPENDIX 2: STATISTICAL CONSIDERATIONS FOR BRII-198 + BRII-196 69
7.1 Randomization Details 69
7.2 Secondary Outcome Measures 69
7.3 Analysis Approaches 69

8 APPENDIX 3: STATISTICAL CONSIDERATIONS FOR AZD7442 IV 70
8.1 Secondary Outcome Measures 70
8.2 Analysis Approaches 70

9 APPENDIX 4: STATISTICAL CONSIDERATIONS FOR AZD7742 IM 71
9.1 Secondary Outcome Measures 71
9.2 Analysis Approaches 71

10 APPENDIX 5: STATISTICAL CONSIDERATIONS FOR SNG001 72
10.1 Objectives 72
10.1.1 Secondary Objectives 72
10.1.2 Exploratory Objectives 73
10.2 Outcome Measures 73
10.2.1 Secondary Outcome Measures 73
10.2.2 Other Outcome Measures 73
10.3 Analysis Approaches 73

10.3.1 Secondary Analyses 73
10.3.2 Exploratory Analyses 74

11 APPENDIX 6: STATISTICAL CONSIDERATIONS FOR CAMOSTAT 75
11.1 Objectives 75
11.1.1 Secondary Objectives 75
11.1.2 Exploratory Objectives 75
11.2 Outcome Measures 75
11.2.1 Secondary Outcome Measures 75
11.3 Analysis Approaches 75

12 APPENDIX 7: STATISTICAL CONSIDERATIONS FOR SAB-185 76
12.1 Study Design and Analysis Population Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185 76
12.2 Additional Analysis Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185 78
12.3 Analysis Considerations for the Non-Omicron Subpopulation Enrolled under Protocol Version 7.0 78
12.4 Data and Safety Monitoring for the Evaluation of SAB-185 under Protocol Version 8.0 78

13 APPENDIX 8: STATISTICAL CONSIDERATIONS FOR BMS-986414 + BMS-986413 79
13.1 Secondary Outcome Measures 79
13.2 Analysis Approaches 79

Version History

Version	Changes Made	Date Finalized
1.0	Original Version	July 29, 2020
2.0	Updated SAP to address the following: - Changes to the protocol based on CMs and LOAs - Typos/errors found after finalization of version 1.0 - Revised handing of missing symptom, virology, and oxygen saturation data in analysis - Clarifying imputation of virology data - Add analyses of resistance mutations - Added LY3819253-specific appendix to address LOA #3	Jan 19, 2021
3.0	New SAP to describe analysis plans for agents introduced in protocol versions 2.0, 3.0 and 4.0	April 2, 2021
4.0	Updated SAP to address the following: - Added appendix for BMS-986414 + BMS-986413 agent introduced in protocol version 5.0 - Added AUC outcome in phase III for SARS-CoV-2 RNA from nasal swabs - Added visit and analysis windows for new study weeks (36, 48, 72) - Clarified aspects of imputation for symptoms duration outcome - Updated plans for determining interim stopping boundaries (using EAST software)	April 15, 2021
5.0	Update SAP to address changed implemented in protocol version 6.0. Specifically: - Updated SAP to reflect changes to study design, randomization, study objectives, interim monitoring, and outcome measures - Clarified that although phase II enrollment was restricted to those at ‘lower’ risk of progression, all participants who enrolled under prior versions of the protocol would be included in all analyses (i.e., ‘higher’ risk participants) - Removed risk stratification subgroup analyses for phase III outcome measures as all phase III is restricted to those at ‘higher risk’	June 24, 2021

	- Clarified that risk stratification subgroup analyses for phase II outcome measures will depend on the number of 'higher' risk participants enrolled - Added supportive and sensitivity analyses for phase II primary symptom outcome measure per FDA recommendation - Removed exploratory virology analyses - Edited typographical errors	
6.0	Update SAP to address changed implemented in protocol version 7.0 and Letter of Amendment 1. Specifically: - Updated SAP to reflect changes in study objectives, schedules for some evaluations, and associated outcome measures. - Updated SAP to reflect changes in interim monitoring, mainly related to changes in graduation criteria. - Updated this SAP to focus on the placebo-controlled phase III evaluation. (Note that a separate *revision to the* SAP will be prepared for the active-controlled phase III evaluation introduced in protocol version 7.0). [Note that the italicized text was added in version 7.0 of the SAP].	September 13, 2021
7.0	Updated SAP with the following major changes: - To add details that are specific to the active-controlled phase 3 evaluation of agents based on protocol version 7.0 and Letter of Amendment 1. - To edit some text to provide clarity concerning the analysis approaches which are the same regardless of whether a placebo control or active control is involved. In part, to achieve this, the terminology "comparator intervention" is often used. - To replace the previous section 5.4 concerning interim analysis considerations for the placebo-controlled phase III trial (which have been completed) with interim analysis considerations for the active-controlled phase III trial. - To indicate exclusion from analysis of viral shedding results from samples labelled as 'Thawed', 'Destroyed', 'Quantity Not Sufficient' or 'Invalid Specimen' as approved by the trial sponsor. - To focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.	October 24, 2021

8.0	Updated SAP to address changes implemented in Letter of Amendment 2 to protocol version 7.0. Specifically: - Added oxygen saturation as a secondary outcome in phase 3 - For the SNG001 agent, added phase II secondary objectives. - For the SNG001 agent, revised phase III exploratory objectives. - For the SNG001 agent, added details on subgroup analyses. Removed Hodges-Lehmann analysis from all analysis sections as the validity of this analysis is questionable for the type of data being generated in this study for the affected outcome measures.	January 24, 2022
9.0	Updated SAP to address changes implemented in protocol version 8.0. Specifically: - Added design details of new superiority placebo-controlled phase III trial that was introduced, including differences from the phase III superiority evaluation of the BRII agent - Outlined statistical analyses of the SAB-185 agent including implications of changing the phase III design due to the Omicron variant - Changed the Primary Symptom Duration Outcome in phase II as per change in protocol version 8.0 - Corrected the phase II “other” outcome measure for the SNG001 agent	February 25, 2022
9.1	Updated SAP to note that LOA #1 to Protocol Version 8.0 has been implemented and no changes to the Primary SAP are needed. Corrected a typo by deleting ‘cough’ from the exploratory analysis approaches section of the SNG001 appendix, as this was removed from the exploratory objective and other outcome in prior SAP versions, but was not deleted from the analyses approach section.	April 12, 2022
10.0	Updates to this version of the SAP pertain to the phase III evaluation of the SAB-185 agent. This evaluation was terminated early based on a DSMB recommendation because of operational futility. Protocol version 8.0 and SAP version 9.0 introduced a plan for the comparison of the SAB-185 agent versus the combination of casirivimab plus imdevimab to be undertaken separately in two subpopulations, referred to as the Omicron	January 9, 2023

	Subpopulation and Non-Omicron Subpopulation. The definitions of these two subpopulations were based on available variant data for participants in the study and, for participants without variant information, on CDC surveillance data concerning the timing of emergence of the Omicron variant. The definitions are updated in this SAP to be based on the timing of emergence of the Omicron variant within the study population. This is to reduce the risk of misclassification between the two subpopulations and hence improve the value of information from the study, particularly for secondary outcomes, despite its early termination for operational futility. The update to these definitions also allowed for variant information from samples obtained during follow-up to be used in the classification to subpopulation rather than just use information from samples obtained prior to randomized treatment being given. Additional updates were made to adjust details of the analysis plan that are no longer relevant because of the termination of enrollment to the Phase III evaluation of SAB-185.	

Glossary of Terms

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AE	Adverse Event
AUC	Area Under the Curve
CM	Clarification Memo
COVID-19	Coronavirus Disease 2019
DSMB	Data and Safety Monitoring Board
ECMO	Extracorporeal Membrane Oxygenation
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
LOA	Letter of Amendment
LoD	Limit of Detection
LLoQ	Lower Limit of Quantification
LTFU	Loss to Follow Up
MCAR	Missing Completely at Random
mITT	Modified Intent-to-Treat
NIAID	National Institute of Allergy and Infectious Diseases
NP	Nasopharyngeal
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOE	Schedule of Evaluations
TOC	Trial Oversight Committee
ULoQ	Upper Limit of Quantification

1 Introduction

1.1 Purpose

This Primary Statistical Analysis Plan (referred to as “SAP” in this document) describes the general framework for the interim and key statistical analyses of the phase II and phase III placebo-controlled investigations of ACTIV-2/A5401, as well as the phase III active-controlled investigation introduced in protocol version 7.0 and the phase III placebo-controlled investigation introduced in protocol version 8.0 . This SAP addresses the primary and secondary objectives and associated outcome measures, as well as a subset of exploratory objectives and associated outcome measures that may be included in primary manuscripts of the study. Hence, it also describes the primary and secondary outcome measures for which results will be posted on ClinicalTrials.gov. This SAP outlines the general statistical approaches that will be used in the analysis of the study and has been developed to facilitate discussion of the statistical analysis components among the study team, industry collaborators, and study sponsor; and to provide agreement between the study team and statisticians regarding the statistical analyses to be performed and presented. Given the design of the study and that, multiple investigational agents will be studied; separate analysis reports may be generated for each investigational agent and each study phase. Analysis considerations that are specific to a given investigational agent are provided in agent-specific appendices to this SAP.

1.2 Version History of this SAP

ACTIV-2 is a platform trial designed to evaluate multiple agents under a master protocol. Versions 1.0 and 2.0 of the SAP, which were based on protocol version 1.0, were developed with the idea that they would be applied to all agents included in the study. However, there were sufficient changes between protocol version 1.0 and subsequent versions of the protocol that versions 1.0 and 2.0 of the SAP were limited to analyses of data evaluating the first agent in ACTIV-2, referred to as LY3819253.

Version 3.0 of the SAP was developed for agents entering under protocol versions 2.0, 3.0 and 4.0, and was not used to describe analyses of data for LY3819253. Because version 3.0 of the SAP applied to different agents from version 2.0 of the SAP, changes between version 2.0 and version 3.0 of the SAP are not detailed here. Analyses that are only for a specific agent or agents are described in agent-specific supplements to the SAP. SAP version 4.0 was developed to address changes in protocol version 5.0 and to make adjustments noted in the version history table.

SAP version 5.0 was developed to address changes made to the protocol in version 6.0. Protocol version 6.0 stated that enrollment to all agents (except BRII-196+BRII-198 which was already in phase III), will stop after the phase II enrollment is completed and there will be no enrollment to a placebo-controlled phase III evaluation of these agents (note that this was subsequently changed in protocol version 8.0—see below). SAP version 5.0 therefore described planned statistical analyses for both the phase II and the phase III evaluations of BRII-196+BRII-198 versus placebo, and for the phase II evaluations of all other agents that entered the study under protocol versions 3.0, 4.0 and 5.0 and which are also being compared to a placebo. In

addition, SAP version 5.0 addressed some small changes to the schedule of evaluations and outcome measures introduced in protocol version 6.0.

Protocol version 7.0 introduced an open-label non-inferiority phase III design to compare investigational agents to an active-comparator among persons at higher risk of progression to hospitalization or death. This phase III evaluation was separate from the phase II superiority evaluation of agents compared to placebo among persons at lower risk for progression to hospitalization or death. Changes introduced in SAP version 6.0 focused on changes made under protocol version 7.0 (and letter of amendment #1) that related to the placebo-controlled superiority phase II/III design (note: BRII-196+BRII-198 is the only agent that enrolled in the placebo-controlled phase III design until protocol version 8.0 was introduced when a placebo-controlled design was also used for the phase III evaluation of SAB-185). Changes introduced in SAP version 7.0 addressed the introduction of the active-controlled non-inferiority phase III trial in protocol version 7.0 (and letter of amendment #1). SAP version 7.0 also introduced the exclusion from statistical analysis of results generated from problematic virologic samples based on a decision made by the DAIDS and study team. In addition, section 5.4 concerning interim analysis considerations was revised to replace considerations for the placebo-controlled phase III trial for which DSMB monitoring had been completed with considerations for the active-controlled phase III trial. Finally, adjustments were made to focus subgroup analysis by country on analyses for participants enrolled at U.S. versus non-U.S. sites, and to add a subgroup analysis by SARS-CoV-2 variants.

SAP version 8.0 implemented changes made under letter of amendment #2 to protocol version 7.0, which added oxygen saturation outcome for the active-controlled phase III and new phase III secondary and exploratory objectives for the SNG001 agent. In addition, analyses using the Hodges-Lehmann estimate were removed throughout as the validity of these analyses is questionable for the type of data being generated in this study for the affected outcome measures.

While the phase III evaluation of SAB-185 was ongoing using an active-controlled non-inferiority design, in vitro data suggested that the active control agent would not have activity against the newly emergent Omicron variant of SARS-CoV-2. As a result, enrollment to the phase III non-inferiority trial was terminated and was redesigned in order to continue a phase III evaluation of SAB-185. This design was defined in protocol version 8.0 and provided for phase III evaluation of an investigational agent versus placebo, but allowing participants to receive other COVID-19 treatments after study entry if available (availability was, however, expected to be very limited). In essence, this led to the reintroduction of a placebo-controlled phase III trial and the general approach for statistical analyses in protocol version 8.0 for this phase III trial follows the earlier plan for the placebo-controlled phase III evaluation of BRII-196+BRII-198. SAP version 9.0 was implemented to describe these changes.

Participants infected with the Omicron SARS-CoV-2 variant who were randomized under protocol version 7.0 to the “active” control agent in the phase III non-inferiority evaluation of SAB-185 were thought to have been treated with an ineffective agent, so functionally with a placebo from an efficacy perspective, The SAB-185-specific appendix of protocol version 8.0 therefore specifies that the subpopulation of participants enrolled in the non-inferiority phase III evaluation of SAB-

185 under protocol version 7.0 who were definitely or very likely infected with the Omicron variant would be included in the analysis population for the placebo-controlled evaluation of SAB-185. This particular nuance is described in more detail in the SAB-185-specific appendix of this SAP.

SAP version 9.0 also included a change to an exploratory objective and associated outcome measure for the evaluation of SNG001 that was introduced in protocol version 7.0 but was not reflected in the applicable previous versions of the SAP.

Updates made in Version 10.0 of the SAP pertain to the phase III evaluation of the SAB-185 agent. This evaluation was terminated early based on a DSMB recommendation because of operational futility. Protocol version 8.0 and SAP version 9.0 introduced a plan for the comparison of the SAB-185 agent versus the combination of casirivimab plus imdevimab to be undertaken separately in two subpopulations, referred to as the Omicron Subpopulation and non-Omicron Subpopulation. The definitions of these two subpopulations were based on available variant data for participants in the study and, for participants without variant information, on CDC surveillance data concerning the timing of emergence of the Omicron variant. The definitions are updated in Version 10.0 of the SAP to be based on the timing of emergence of the Omicron variant within the study population. This is to reduce the risk of misclassification between the two subpopulations and hence improve the value of information from the study particularly for secondary outcomes despite its early termination for operational futility. The update to these definitions also allowed for variant information from samples obtained during follow-up to be used in the classification to subpopulation rather than just use information from samples obtained prior to randomized treatment being given. Additional updates were made to adjust details of the analysis plan that are no longer relevant because of the termination of enrollment to the Phase III evaluation of SAB-185

2 Study Overview

2.1 Study Design

The study design described in this section reflects details in protocol version 8.0 for the placebo-controlled phase II and phase III evaluations of investigational agents. This section also includes a description of the non-inferiority phase III evaluation of investigational agents per protocol version 7.0 and letter of amendments 1 and 2.

ACTIV-2/A5401 is a master protocol to evaluate the safety and efficacy of investigational agents for the treatment of symptomatic non-hospitalized adults with COVID-19. The study is designed to evaluate both infused and non-infused investigational agents.

The trial has a randomized controlled adaptive platform study design that allows agents to be added or dropped during the course of the study for efficient phase II and phase III testing of new agents within the same trial infrastructure.

Version 8.0 of the protocol provides for a blinded phase II evaluation of an investigational agent compared to placebo among participants at lower risk for progression to hospitalization or death, regardless of the mode of administration of the agent; for some agents, enrollment to higher risk participants in phase II was allowed under earlier protocol versions.

Based on protocol-specified criteria, agents could graduate from phase II to phase III evaluation. The phase III evaluation of investigational agents under all protocol versions has been in a population of participants at higher risk of hospitalization or death (though the definition of "higher risk" as changed across protocol versions). When protocol version 8.0 was introduced, only two agents had graduated to phase III evaluation and started enrollment: BRII-196+BRII-198 and SAB-185. For these two agents, protocol version 8.0 provides for:

- Continued follow-up of participants enrolled under protocol versions 2.0 to 6.0 into a placebo-controlled phase III trial evaluating the combination monoclonal antibody agent, BRII-196 + BRII-198.
- Continued follow-up of participants enrolled under protocol version 7.0 into a non-inferiority phase III trial evaluating the polyclonal antibody agent, SAB-185 using the monoclonal antibody combination of casirivimab plus imdevimab (REGEN-COV, Regeneron) as the control regimen. As noted above, enrollment to this non-inferiority trial was terminated because of an anticipated lack of efficacy of casirivimab plus imdevimab against the Omicron SARS-CoV-2 variant that became widely prevalent.
- Enrollment of participants into a placebo-controlled phase III trial evaluating SAB-185. In the latter trial, use of COVID-19 treatments obtained outside of the trial is allowed in both randomized arms, if available (though availability is expected to be very limited). Protocol version 8.0 allowed for the inclusion of participants enrolled under protocol version 7.0 who were, or were very likely, infected with the Omicron variant to be included in this evaluation as it was thought that casirivimab plus imdevimab would be ineffective against this variant (and hence participants receiving this combination could be considered as functionally treated with a placebo). Enrollment to this trial was terminated early based on a recommendation from the DSMB because of operational futility, specifically a very low rate of hospitalizations/deaths. This was done before randomization to a placebo control group was started, so the evaluation of SAB-185 in the Omicron Subpopulation is actually versus a control group of participants who were randomized to casirivimab plus imdevimab

When two or more agents are being evaluated in the same phase of the study, the trial design includes sharing of the appropriate control group (placebo or active comparator) for efficient evaluation of each agent. Note, however, that enrollment to the phase III placebo-controlled evaluation of BRII-196+BRII-198 did not coincide with enrollment the phase III placebo-controlled evaluation of SAB-185 and so there is no sharing of the placebo control group for these two agents.

Eligible participants enrolled under all versions of the protocol from version 2.0 have intensive follow-up through day 28, followed by limited follow up through at least week 72 weeks in phase II and phase III.

The study population consists of adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry (this criterion allowed up to 10 days in

protocol version 3.0 and earlier, and up to 8 days in protocol versions 4.0 and 5.0), and with presence of select symptoms within 24 hours of study entry.

2.2 Randomization Process

Under protocol version 8.0, the phase II trial and the phase III trial involve different populations and have separate randomizations. However, the structure of the randomization process is the same for each of the two trials, as described in the following.

The randomization process is designed to be flexible for this adaptive platform study, in which participants may be eligible for randomization to different investigational agents, and investigational agents can be added or dropped during the course of the study. The ultimate intent is to have a similar number of concurrently randomized participants on a given investigational agent and in the placebo comparator group for that agent (i.e. combining participants who were eligible to receive the agent but who were randomized to any of the available placebos for investigational agents in the same phase of evaluation).

To achieve having a similar number of participants on the active arm and in the pooled comparator group for a given investigational agent, the randomization occurs in two steps within each trial.

The first randomization is to *Agent Group*. For a given participant, the first randomization assigns a participant with equal probability among the n agents in the trial (e.g., a 1:1 ratio for two agents, 1:1:1 ratio for three agents, etc.) that the participant is eligible to receive (based on protocol eligibility criteria and the set of agents available at the clinical site at which the participant is being enrolled). Trial phase for an agent is accounted for in the participant eligibility (i.e. by the classification of their risk for hospitalization/death as lower or higher). In the event that a participant is only eligible for one investigational agent (n=1), then they are assigned to the one appropriate Agent Group.

Immediately following the first randomization, participants are randomized within an Agent Group in the second randomization to the investigational agent or appropriate comparator (i.e., the matching placebo for agents in the same phase of evaluation). For a given participant, the probability of assignment to the investigational agent or placebo in the second randomization depends on the number of agents currently under investigation that the participant was eligible to receive, as phase II and phase III have distinct populations (phase II is restricted to those at lower risk of progression to hospitalization and death, and phase III is restricted to those at higher risk).

Both the first and second randomizations involve blocked stratified randomization. In phase II, both the first and second randomizations are stratified by time from symptom onset (≤5 days vs >5 days), however, in previous versions of the protocol, in which both ‘higher’ and ‘lower’ risk participants could be randomized to agents in phase II evaluation, both the first and second randomizations were also stratified by risk group (‘higher’ vs ‘lower’). In the active-controlled phase III trial introduced in protocol version 7.0 and the placebo-controlled trial introduced in protocol version 8.0, both randomization steps are stratified by country. Under previous versions of the protocol for the placebo-controlled phase III trial evaluating BRII-196+BRII-198, both

randomization steps were only stratified by time from symptom onset (≤ 5 days vs > 5 days). Additional details on randomization are provided in protocol section 10.3.

2.3 Study Objectives

The following sections list the primary, secondary and exploratory objectives from protocol version 8.0; corresponding protocol numbering is shown in brackets. This Primary SAP addresses all of the primary and secondary objectives shown below, with the exception of the secondary PK objectives in phase II, which will be addressed outside of this SAP. In addition, exploratory objectives 1 and 4 will also be addressed in this SAP; however, other exploratory objectives will be addressed separately.

2.3.1 Primary Objectives

1) Phases II and III: To evaluate safety of the investigational agent [Protocol Objective 1.1.1].

2) Phase II: To determine efficacy of the investigational agent to reduce the duration of COVID-19 symptoms through study day 28 [Protocol Objective 1.1.2].

3) Phase II: To determine the efficacy of the investigational agent to increase the proportion of participants with nasopharyngeal (NP) SARS-CoV-2 RNA below the lower limit of quantification (LLoQ) at study days 3, 7, and 14 [Protocol Objective 1.1.3].

4) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28. Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic [Protocol Objective 1.1.4].

2.3.2 Secondary Objectives

1) Phases II and III: To determine whether the investigational agent reduces a COVID-19 severity ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 [Protocol Objective 1.2.1].

2) Phases II and III: To determine whether the investigational agent reduces the progression of COVID-19-associated symptoms [Protocol Objective 1.2.2].

3) Phases II and III: To determine if the investigational agent reduces levels of SARS-CoV-2 RNA in NP swabs [Protocol Objective 1.2.3].

4) Phase III: To determine the efficacy of the investigational agent to increase the proportion of participants with NP SARS-CoV-2 RNA below the LLoQ at study day 3 [Protocol Objective 1.2.4]

5) Phase II: To determine the pharmacokinetics of the investigational agent [Protocol Objective 1.2.5].

6) Phase II and III: To determine efficacy of the investigational agent to obtain pulse oximetry measurement of ≥ 96% through day 28 [Protocol Objective 1.2.6].

7) Phase III: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72 [Protocol Objective 1.2.7].

8) Phase III: To evaluate if the investigational agent reduces the time to sustained symptom resolution through study day 28 [Protocol Objective 1.2.8].

9) Phase III: To determine if the investigational agent will prevent the composite endpoint of hospitalization or death through study day 28, excluding hospitalizations that are determined to be unrelated to COVID-19 [Protocol Objective 1.2.9].

2.3.3 Exploratory Objectives

1) Phases II and III: To explore the impact of the investigational agent on participant-reported rates of SARS-CoV-2 positivity of household contacts [Protocol Objective 1.3.1].

2) Phases II and III: To explore if baseline and follow-up hematology, chemistry, coagulation, viral, and inflammatory biomarkers are associated with clinical and virologic outcomes in relation to investigational agent use [Protocol Objective 1.3.2].

3) Phases II and III: To explore possible predictors of outcomes and differences between investigational agent and control (placebo in phase II and active comparator in phase III) across the study population, notably sex, time from symptom onset to start of investigational agent, and race/ethnicity [Protocol Objective 1.3.3].

4) Phases II and III: To explore if the investigational agent changes the hospital course in those hospitalized [Protocol Objective 1.3.4].

5) Phases II and III: To explore and develop a model for the interrelationships between virologic outcomes, clinical symptoms, and, in phase III, hospitalization, and death in each study group [Protocol Objective 1.3.5].

6) Phases II and III: To explore the relationship between exposure to the investigational agent and SARS-CoV-2 innate, humoral or cellular response, including anti-drug antibodies, as appropriate per investigational agent [Protocol Objective 1.3.6].

7) Phases II and III: To explore baseline and emergent viral resistance to the investigational agent [Protocol Objective 1.3.7].

8) Phases II and III: To explore the association between viral genotypes and phenotypes, and clinical outcomes and response to agents [Protocol Objective 1.3.8].

9) Phases II and III: To explore the association between host genetics and clinical outcomes and response to agents [Protocol Objective 1.3.9]

10) Phases II and III: To explore relationships between dose and concentration of investigational agent with virology, symptoms, and oxygenation [Protocol Objective 1.3.10].

11) Phases II and III: To explore the prevalence, severity, and types of persistent symptoms and clinical sequelae in participants through end of study follow-up [Protocol Objective 1.3.11].

12) Phases II and III: To explore measures of psychological health, functional health, and health-related quality of life in participants through end of study follow-up [Protocol Objective 1.3.12].

2.4 Overview of Sample Size Considerations

The sample size for phase II was the same under protocol versions 2.0 to 8.0. The sample size for the placebo-controlled superiority phase III design was also the same under protocol versions 2.0 to 6.0 (it was originally defined in Appendix IV of protocol version 2.0 for the BRII-196+BRII-198 agent entered into the study under protocol version 2.0) and is currently detailed in Appendix V of protocol version 8.0 for the BRII-196+BRII-198 agent. The sample size for the non-inferiority phase III design that enrolled participants under protocol version 7.0 is detailed in section 10.4 of protocol version 7.0. As noted above, enrollment to that non-inferiority trial was terminated early. The sample size for the placebo-controlled phase III trial introduced in protocol version 8.0 is detailed in section 10.4 of protocol version 8.0.

2.4.1 Phase II – Placebo-Controlled Superiority

For each investigational agent in phase II, the proposed sample size is 220 participants, consisting of 110 participants who receive that agent and 110 participants who are concurrently randomized to placebo control. Participants who are randomized but do not start their randomized investigational agent or placebo will not be followed.

This sample size is chosen to give high power to identify an active agent on the basis of the primary virology outcome, due to limited data on the variability of symptom duration in the outpatient COVID-19 population.

Assuming 100 participants in each group will have NP swabs available at a scheduled measurement time, there is at least 82% power to detect a 20% absolute increase in the percentage of participants with SARS-CoV-2 RNA < LLoQ in the investigational agent group vs concurrent placebo group, regardless of the assumed percent <LLoQ in the placebo group

(range: 10-70%); calculated for the comparison of two proportions using a normal approximation to the binomial distribution, unpooled variance, and two-sided Type I error rate of 5%.

2.4.2 Phase III – Placebo-Controlled Superiority Trial Used for the Evaluation of BRII-196+BRII-198

The proposed sample size was 842 participants consisting of 421 participants who received the active agent and 421 participants who were concurrently randomized to placebo control. This sample size included the enrollment that occurred during the phase II placebo-controlled evaluation of an agent. Participants who were randomized but did not start their randomized investigational agent or placebo were not followed.

This sample size was chosen to provide 90% power to detect a relative reduction of 50% in the proportion of participants hospitalized/dying between the study groups. This was based on the following assumptions:

- Proportion hospitalized/dying in the placebo group is 15%;
- Two-sided test of two proportions with 5% Type I error rate;
- Three interim analyses and one final analysis, approximately equally spaced, with stopping guideline for efficacy of an investigational agent versus concurrent placebo determined using the Lan-DeMets spending function approach with an O'Brien and Fleming boundary, and a non-binding stopping guidelines for futility using a Gamma(-2) Type II spending function also implemented using the Lan-DeMets spending function;
- Allowance for 5% of participants to be lost-to-follow-up prior to being hospitalized or dying, and non-informative loss-to-follow-up.

2.4.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant)

The active-controlled Phase III trial was focused on a non-inferiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus an active comparator agent, specifically the monoclonal antibody combination of casirivimab plus imdevimab. The non-inferiority margin for the absolute difference in proportion hospitalized/dead was 3% (investigational agent minus active comparator agent); the rationale for this choice was described in Section 3.1 of protocol version 7.0. Non-inferiority was considered to be established if a two-sided exact 95% confidence interval for the absolute difference was entirely below 3%. Details of the construction of the confidence interval are in section 10.6 of protocol version 7.0 and are included further below in this SAP.

The sample size differed between infused investigational agents (600 for the investigational agent and 600 for the concurrently randomized active comparator) and non-infused investigational agents (800 per arm instead of 600 per arm). The rationale for this was that there may be broader clinical utility for non-infused agents such that a slightly higher true hospitalization/death rate may be tolerated in clinical practice. No enrollment occurred for a non-infused agent and so the

sample size justification described below is just for an infused agent (enrollment only occurred for the SAB-185 infused agent).

Sample Size Justification for Infused Investigational Agents

For the evaluation of a specific infused investigational agent, the sample size was 1200 participants including approximately 600 participants randomized to receive the infused investigational agent and approximately 600 participants (who were eligible to receive the infused investigational agent) concurrently randomized to receive the active comparator agent. This sample size was chosen to provide close to 90% power to establish non-inferiority assuming that the true proportion hospitalized/dead for both the infused investigational agent and the active comparator agent was 2.3%. The rate of 2.3% was based on the observed proportion for casirivimab plus imdevimab combining across doses in the subpopulation of the Regeneron COV-2067 clinical trial who met the criteria for being at high risk of progression to hospitalization/death (FDA communication to DAIDS/NIAID, May 2021). No adjustment for loss to follow-up was made in the sample size as the primary analysis was to be based on the observed number of hospitalizations divided by the number of participants who initiated study treatment. In addition, the impact of any loss to follow-up was expected to be minimal as there was to be regular contact between research site staff and participants (or their secondary contacts) and previous experience in the study and other trials has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up).

The potential power of the study was evaluated in two ways using the PASS version 15 sample size calculation software. Both used a non-inferiority hypothesis testing approach based on use of the Miettinen and Nurminen score test statistic (which was the basis for calculating the confidence interval used for the analysis). The first ignored interim monitoring but used a binomial enumeration method to calculate power and type I error rates. Use of the binomial enumeration method takes account of the discreteness of the binomial distribution (rather than using a normal approximation to the binomial distribution) which may be important in the setting of low hospitalization/death probabilities. Using this approach gave a power of 90.2%. The second approach did not use a binomial enumeration but took account of interim analyses using a standard implementation of four equally-spaced interim analyses using the O'Brien and Fleming stopping guideline. This used a simulation approach and gave a power of 90.0% (width of 95% confidence interval around this simulation-based value was 0.12%). Based on these two approaches, it was anticipated that the study would have had close to 90% power to show non-inferiority for an infused investigational agent assuming that it truly had the same 2.3% hospitalization/death rate as the active comparator agent.

The PASS software was also used to illustrate how the power of the study might change for various scenarios which differed from the scenario assumed (see Table below). This was undertaken using the first of the two approaches mentioned above (i.e., using the binomial enumeration approach). Looking at the top part of the table in which both the infused investigational agent and the active comparator agent have the same underlying true hospitalization/ death rate, the power is decreased if the true rate was above the assumed 2.3%, but increased if the true rate was less than 2.3%. If the true rate was 3%, then the power was still above 80%, but if the true rate is 4% it was reduced to 73%.

The middle and lower parts of the table show scenarios in which the infused investigational agent had a true hospitalization/death rate of 0.5% or 1% worse than the active comparator agent, respectively. If the true rate for the active comparator agent was 2.3% and was 2.8% for the infused investigational agent (i.e., 0.5% worse), then the power was reduced to 73%. If the true rate for the active comparator agent was 2.3% and was 3.3% for the infused investigational agent (i.e., 1% worse), then the power was reduced to 50%

Table: Power for various scenarios based on non-inferiority hypothesis testing using the likelihood score test statistic (Miettinen and Nurminen method) with binomial enumeration of power and Type I error rate. All scenarios use a 3% non-inferiority margin and one-sided Type-I error rate of 0.025 with a sample size of 600 participants receiving an infused investigational agent and 600 participants receiving the active comparator agent. Power in practice would have been slightly reduced from the values shown due to interim monitoring.

Same Underlying True Hospitalization/Death Rate For Active Comparator Agent and Infused Investigational Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
99.4%	1%	1%	2.2%
97.3%	1.5%	1.5%	2.2%
93.2%	2%	2%	2.3%
90.2%	2.3%	2.3%	2.4%
88.1%	2.5%	2.5%	2.4%
83.1%	3%	3%	2.4%
78.1%	3.5%	3.5%	2.4%
73.2%	4%	4%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 0.5% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
92.5%	1%	1.5%	2.2%
85.2%	1.5%	2%	2.2%
77.4%	2%	2.5%	2.3%
73.1%	2.3%	2.8%	2.4%
70.5%	2.5%	3%	2.4%
64.8%	3%	3.5%	2.4%
59.6%	3.5%	4%	2.4%
55.0%	4%	4.5%	2.4%
Infused Investigational Agent with Underlying True Hospitalization/Death Rate that is 1% Worse than Active Comparator Agent			
Power	True % Hosp/Died on Active Comparator Agent	True % Hosp/Died on Infused Investigational Agent	Actual Type I Error Rate*
71.3%	1%	2%	2.2%
61.7%	1.5%	2.5%	2.2%
54.0%	2%	3%	2.3%
50.4%	2.3%	3.3%	2.4%
48.4%	2.5%	3.5%	2.4%
44.0%	3%	4%	2.4%
40.0%	3.5%	4.5%	2.4%
36.7%	4%	5%	2.4%
*Actual type I error rate is slightly lower than assumed rate of 2.5% because of discreteness of the binomial distribution.			

2.4.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 (Terminated Early Based on the Recommendation of the DSMB Due to Operational Futility)

The proposed sample size is 1200 participants consisting of approximately 600 participants who are randomized to receive the active agent and approximately 600 participants who are concurrently randomized to placebo control. Unlike the placebo-controlled phase III evaluation of investigational agents under earlier versions of the protocol, under protocol version 8.0, participants enrolled in the phase II evaluation of an investigational agent are not part of the study population for the phase III evaluation of the same agent (as participants in the phase II evaluation were generally “lower risk” and participants in the phase III evaluation are “higher risk” for hospitalization/death). Participants who are randomized but do not start their randomized investigational agent or placebo are not followed.

The phase III trial is focused on a superiority comparison of the proportion of participants who are hospitalized or who die through to 28 days for an investigational agent versus placebo (with use of SOC treatment in both arms, if available). The primary analysis will focus on evaluating the ratio of proportions (investigational agent/placebo) or, equivalently, the relative reduction in risk of hospitalization/death for the active investigational agent versus placebo. The sample size of 1200 participants, with approximately 600 randomized to an investigational agent and 600 to placebo, has been chosen to give good power (>90%) to detect relative risk reductions of 70% (as found for other antibody treatments) if the proportion hospitalized/dead in the placebo group is about 5% or higher, using a two-sided Type I error rate of 5%. There are multiple factors that will affect the power, which are discussed below. To provide context for this discussion, the table below shows the power of the study to detect relative risk reductions of between 50% and 70% for proportions hospitalized/dead in the placebo group of 3% to 6%. The powers shown were obtained in PASS software (version 15.0.4) for testing two proportions using a z-test (so the normal approximation method) with unpooled variance. They are based on an effective sample size of 570 per arm, with the 5% reduction from 600 per arm built in to allow for loss to follow-up and interim monitoring using the O’Brien and Fleming stopping guideline.

Power to detect various true effect sizes (relative reduction in risk of hospitalization/death) for selected true proportions hospitalized/dead on placebo between 3% and 6%			
Proportion Hospitalized/Dead		Relative Risk Reduction for Active versus Placebo	Power
Placebo	Active		
3%	0.9%	70%	73%
	1.2%	60%	56%
	1.5%	50%	40%
4%	1.2%	70%	85%
	1.6%	60%	69%
	2.0%	50%	51%
5%	1.5%	70%	92%
	2.0%	60%	79%
	2.5%	50%	61%
6%	1.8%	70%	96%
	2.4%	60%	86%
	3.0%	50%	69%

Discussion of Factors Affecting the Power of the Study

a. Proportion hospitalized/dead in placebo control group: As can be seen in the table, the proportion hospitalized/dead in the placebo arm has a reasonable effect on power with lower proportions leading to a reduction in power. For the placebo control group for evaluating the BRII agent in ACTIV-2, the proportion was 11% [25]. However, the proportion in the phase 3 trial of sotrovimab was 6% [26]. There is also a possibility that the proportion may be lower for the Omicron variant than with previous variants.

b. Use of SOC treatment by some participants: Higher use of SOC treatment will reduce the proportion hospitalized/dead in both randomized arms. For example, the proportion hospitalized/dead in the placebo arm would change from 6% if none receive SOC treatment to 5.58%, 4.74% and 3.90% if SOC treatment is used by a random sample of 10%, 30% and 50%, respectively, of participants in the placebo arm (i.e., SOC treatment use is not related to risk of hospitalization/death) and SOC treatment reduces risk of hospitalization/death by

70%. If SOC treatment use is not random, for example, it is taken up by the highest risk participants, then the impact might be larger. As the trial excludes participants who have accessed SOC treatment prior to entry and there is a general lack of availability of such treatments globally, use in the trial is expected to be very low (e.g., <10%) and so will limit the impact.

c. Differential effect of an investigational agent versus placebo according to use or not of SOC treatment: For a given proportion of participants hospitalized/dead in the placebo arm, the power shown in the above table is valid if the relative effect of SAB versus placebo is not affected by the use of SOC treatment. Power would be reduced from the values shown if the effect of SAB versus placebo is reduced in the presence versus absence of background therapy. A related concern arises if use of SOC treatment is differential in the investigational agent arm versus the placebo arm. For example, accessing SOC treatment at a higher rate in the placebo arm because more participants have a deteriorating health status might diminish a true difference in effect between arms and hence reduce power. As noted above, use of SOC treatment is expected to be low and so any reduction in power is expected to be limited even if this occurs.

d. Failure to start randomized treatment and loss to follow-up: The impact of any loss to follow-up is expected to be minimal as there will be regular contact between research site staff and participants (or their secondary contacts). Previous experience in the study and other trials has shown that the large majority of hospitalizations/deaths occur early in follow-up (first two weeks of follow-up) when loss to follow-up has also been minimal (approximately 1 to 1.5%) in this study. In addition, a very small proportion of randomized participants will not start study treatment and will be excluded from the analysis of the primary outcome. Based on ACTIV-2 experience, allowance for 3-4% not starting treatment or being lost to follow-up before hospitalization is built into the above power table (with additional allowance of 1-2% for interim monitoring using the O'Brien and Fleming stopping guideline).

Because of these uncertainties, the DSMB will be asked to monitor the potential impact of the above factors on the operational feasibility of the study.

2.5 Overview of Formal Interim Monitoring

During the course of the study (phase II and phase III), an independent NIAID-appointed Data and Safety Monitoring Board (DSMB) will undertake reviews of interim data from the study. The following sections outline plans for interim monitoring of the placebo-controlled superiority phase II, the placebo-controlled superiority phase III for evaluating BRII-196+BRII-198, the active-controlled non-inferiority phase III for evaluating SAB-185 under protocol version 7.0, and the placebo-controlled superiority phase III for monitoring investigational agents under protocol version 8.0. Additional details on phase II monitoring can be found in protocol version 8.0 section 10.5.1, and in protocol version 8.0 Appendix V for placebo-controlled phase III monitoring for the BRII-196+BRII-198 agent. Details on active-controlled non-inferiority phase III monitoring for SAB-185 are taken from protocol version 7.0 section 10.5.2 as amended in letter of amendment 1 to protocol version 7.0. Details on the placebo-controlled superiority phase III monitoring introduced under protocol

version 8.0 are taken from section 10.5.2 of protocol version 8.0. Statistical considerations for interim monitoring are described in section 5.4 of this SAP.

Regardless of study phase, in the event that there is any death deemed related to study product or if two participants experience a Grade 4 AE deemed related to study product, enrollment to the study product group will be paused and the DSMB will review interim safety data.

2.5.1 Phase II – Placebo-Controlled Superiority

During phase II, the DSMB will review interim data to ensure the safety of participants in the study, and to evaluate the activity of each investigational agent in order to provide graduation recommendations to the Trial Oversight Committee (TOC) via NIAID. The DSMB may recommend early termination of randomization to a particular investigational agent if there are safety concerns, but it is not intended to stop for futility in the phase II evaluation period.

For each investigational agent, there will be interim analyses of safety data by the DSMB approximately monthly (or on a schedule recommended by the DSMB) with the first review occurring approximately 6 weeks after enrollment to a given agent begins.

2.5.2 Phase III – Placebo-Controlled Superiority Used for the Evaluation of BRII-196+BRII-198

[At the time of preparing this version of the SAP, all participants randomized to receive BRII-196+BRII-198 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study].

During phase III, the DSMB will review interim data to help ensure the safety of participants in the study, and to recommend changes to the study. The DSMB may recommend termination or modification of the study for safety reasons, if there is persuasive evidence of efficacy or lack of efficacy of an investigational agent versus placebo in preventing hospitalizations and deaths, or on the basis of statistical or operational futility. At each interim review, the DSMB will review summaries of data by unblinded randomized arms for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of the investigational agent).

For monitoring the primary efficacy outcome, the O'Brien Fleming boundary will be used as the stopping guideline, implemented using the Lan-DeMets spending function to allow for changes in the timing or number of interim analyses if recommended by the DSMB.

Three interim efficacy analyses are planned during phase III. The first review is planned at the completion of day 28 of follow-up for phase II participants, and second and third reviews are planned for after about 50% and 75% of the expected maximal efficacy (hospitalization/death) information.

The expected maximal efficacy information available at the planned interim analyses is approximately proportional to the expected number of hospitalizations/deaths under design

assumption parameters. Assuming 15% of participants will be hospitalized/die in the placebo control group and 7.5% will be hospitalized/die in the investigational agent group (i.e., relative reduction of 50%), with 421 participants per group, this corresponds to 95 participants hospitalized/died across both groups. Because of the uncertainty around the design assumptions, interim efficacy analyses will occur as follows (unless DSMB recommends otherwise):

- The first interim analysis for phase III will be when 220 participants from the two groups combined have been followed for the primary outcome assessed at day 28 (this will likely then be the same hospitalization/death information used in the phase II graduation analysis), or when approximately 24 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 421 participants from the two groups combined (50% of the 842) have been followed for the primary outcome assessed at day 28, or when approximately 48 participants in the two groups combined have been hospitalized or have died;
- The earlier of when approximately 632 participants from the two groups combined have been followed for the primary outcome assessed at day 28, or when approximately 72 participants in the two groups combined have been hospitalized of have died.

In considering possible modifications to the study or termination of the study for efficacy, the DSMB may also consider interim results for the secondary outcome of death. The DSMB may make recommendations based on a high level of evidence for a difference between randomized arms, which might be based on application of the O'Brien and Fleming stopping guideline to the death outcome. In these circumstances, consideration should be given to the increased risk of a Type I error.

There is the possibility that differences between the randomized arms may be observed at an early study time point (for example, cumulative proportion at day 6); however, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the randomized arm comparisons will be at day 28.

The DSMB will monitor for statistical futility (i.e., stopping early for the absence of difference between groups). An investigational agent may be discontinued based on evidence of lack of effect or very limited effect compared with placebo control. For the purpose of evaluating statistical futility, a moderately aggressive Type II error spending function, Gamma (-2) spending function implemented using the Lan-DeMets spending function approach, will be used.

The DSMB will also monitor operational futility. With respect to operational futility, the DSMB may recommend modification or termination of the study if the proportion hospitalized/die in the control group is much lower than expected in designing the trial. For example, the DSMB might recommend restricting or closing enrollment to the low-risk stratum in favor or increasing enrollment to the high-risk stratum. In addition, the DSMB will monitor the loss to follow-up (LTFU) rate. As a benchmark, an overall LTFU rate of more than 10% would be cause for concern.

2.5.3 Phase III – Active-Controlled Non-Inferiority Trial Used for the Evaluation of SAB-185 under Protocol Version 7.0 (Terminated Early Because the Control Regimen Was Considered Ineffective Against the Predominant SARS-COV-2 Omicron Variant)

[At the time of preparing this version of the SAP, all participants randomized in the non-inferiority phase III evaluation of SAB-185 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study. Note that the first interim analysis for this part of the study was scheduled before the termination of enrollment due to the anticipated lack of efficacy on the control regimen against the SARS-CoV-2 Omicron variant that had become widely prevalent. That interim analysis was replaced by an interim analysis that followed the monitoring approach described in Section 2.5.4 and the SAB-185-specific appendix of this SAP].

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if there is persuasive evidence of non-inferiority (or superiority or inferiority) of an investigational agent versus the active comparator agent in its effect on the hospitalization/death outcome. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend termination or modification of the study if it appears futile on statistical or operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, three interim analyses for DSMB review are planned for each investigational agent, after approximately 25%, 50% and 75% of the planned enrollment for an investigational agent has been completed and followed through to day 28. At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guidelines for Efficacy or Lack of Efficacy

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses) for the absolute difference (investigational agent minus active comparator agent) in the proportion of participants hospitalized or dead by day 28, relative to thresholds defining non-inferiority, superiority or inferiority of the investigational agent as follows (in the order given):

- The DSMB may recommend releasing results evaluating the effect of an investigational agent when both non-inferiority and superiority of that agent is established based on the confidence interval being entirely below 0% (i.e., supportive of a lower true proportion being hospitalized or dying on the investigational agent than the active comparator agent). If this occurs, consideration will need to be given to the ongoing appropriateness of the active comparator agent as a control for evaluating other investigational agents in the study.

- Early stopping and/or release of results based on non-inferiority should be considered on an agent-by-agent basis. For non-infused agents, the DSMB may recommend releasing results evaluating the effect of an investigational agent when non-inferiority (but not superiority) of that agent is established based on the confidence interval being entirely below 3% (but not entirely below 0%). However, in the interests of also having an adequate safety database for the investigational agent, it is not intended that this recommendation be made before approximately 400 participants have been randomized to receive the agent (or some other number of participants specified in the agent-specific appendix). In addition, the study may continue randomizing participants to the investigational agent in the interests of increasing precision in evaluating the agent; this decision will be made by the study team and sponsor on an agent-by-agent basis. For infused agents, early stopping and/or release of results for non-inferiority should not be considered.

- The DSMB may recommend releasing results and terminating randomization to an investigational agent if inferiority of that agent is established based on the confidence interval being entirely above 0% (i.e., suggesting a higher true proportion being hospitalized or dying on the investigational agent than the active comparator agent). Examples of how this criterion might be met when evaluating an infused agent and when the observed control rate is close to 2.3% include observing 18/150 versus 3/150 (observed difference 10.0%) at the first interim analysis; 23/300 versus 7/300 (observed difference 5.3%) at the second interim analysis; and 24/450 versus 10/450 at the third interim analysis (observed difference 3.1%). In these examples, all observed differences are higher than the non-inferiority margin of 3%, and are indicative also of the futility of continuing evaluation of the infused investigational agent to demonstrate non-inferiority.

2.5.4 Phase III – Placebo-Controlled Superiority Trial Introduced Under Protocol Version 8.0 (Terminated Early Based on the Recommendation of the DSMB Due to Operational Futility)

[At the time of preparing this version of the SAP, all participants randomized in the superiority phase III evaluation of SAB-185 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study].

The DSMB will undertake reviews of interim data from the study to help ensure the safety of participants in the study, and to recommend changes to the study including termination or modification for safety reasons or if efficacy of the agent versus placebo has been established, or if it is unlikely that the agent has sufficient efficacy to warrant further evaluation in this study. It is not intended, however, to terminate evaluation of an agent early for efficacy based on symptom outcome measures. The DSMB may also recommend termination or modification of the study if it appears futile on operational grounds to continue an investigational agent in the study as designed.

Unless otherwise recommended by the DSMB, two interim analyses for DSMB review are planned for each investigational agent, after approximately one-third (i.e., approximately 400

participants) and two-thirds (i.e., approximately 800 participants) of the planned enrollment for an investigational agent has been completed and followed through to day 14 (the choice of day 14 is because the large majority of hospitalizations/deaths in ACTIV-2 have been observed to occur by day 14).

At each interim review of an investigational agent, the DSMB will review summaries of data by randomized treatment arm for the primary outcome of hospitalization/death, the secondary outcome of death, losses to follow-up, and adverse events (including early discontinuation of investigational agent).

Decision Guideline for Efficacy Favoring an Investigational Agent versus Placebo

The general approach for decision-making with respect to efficacy is based on evaluating a two-sided 95% confidence interval (adjusted for interim analyses—see further below) for the relative difference (investigational agent / placebo) in the proportion of participants hospitalized or dead by day 28. As a stopping guideline for greater efficacy of an investigational agent compared with placebo, the O'Brien and Fleming boundary will be used. The stopping guideline will be implemented using the Lan-DeMets spending function approach to allow for the possibility of changes in the timing of interim analyses and/or additional (or fewer) interim analyses if recommended by the DSMB. Information time for the spending function will be based on the proportion of the planned enrollment (i.e., of the 1200 participants for comparing an investigational agent to placebo) who could have been followed through day 14 at the time of the data freeze for the interim analysis. The choice of day 14 here reflects the fact that the very large majority of hospitalizations and deaths in ACTIV-2 have occurred by 14 days of follow-up. As a guideline, if the two-sided 95% confidence interval (adjusted for interim analyses) excludes a risk ratio of one (equivalently a relative risk reduction of zero) favoring the investigational agent, then the DSMB may recommend closure of randomization to that agent; release of interim results may also be recommended.

There is the possibility that differences between the treatment groups may be observed early in follow-up. However, the overall goal of the study is to prevent hospitalization and deaths regardless of the timing, and therefore the focus of the treatment group comparisons will be on the cumulative proportion hospitalized/dead at day 28.

Stopping Randomization to an Investigational Agent Because of Limited Efficacy

Because there are treatments available that may substantially reduce the risk of hospitalization/death, albeit with limited availability and the caveat that they have generally been evaluated among individuals infected with earlier variants of SARS-CoV-2, it is likely that a treatment which reduces the risk of hospitalization/death by less than 30% versus placebo will have limited utility in clinical practice. Therefore, as a non-binding guideline, the DSMB may recommend early termination of randomization to a specific investigational agent because of limited efficacy if the two-sided 95% confidence interval (adjusted for interim analyses) for the risk ratio is entirely above 0.7 or, equivalently, the two-sided 95% confidence interval (adjusted for interim analyses) for the relative risk reduction is entirely below 30%.

Modifying or Stopping the Study for Operational Futility

The DSMB will also monitor operational futility, in particular related to losses to follow up, low hospitalization/death rate in the placebo arm (which, in part, may arise due to more extensive use of SOC treatment than anticipated). As most hospitalizations are expected to occur early in follow up (e.g., during the first 14 days), early losses to follow up would be most relevant. As a benchmark, an overall loss to follow-up rate (excluding losses after a participant is hospitalized) of more than 5% would be cause for concern.

With regard to the hospitalization/death rate in the placebo arm, the power of the study is limited if this rate is below 3% (see power analysis table above). Therefore, as a benchmark, an observed rate of less than 3% in the placebo arm would be a cause for concern. If this arises, or temporal trends in hospitalization/death rate suggest it might, then any DSMB recommendation concerning this issue might incorporate information about factors that might be driving it (e.g., increasing use of SOC treatment, evolving lower risk of participants enrolled, or lower risk with new variants).

2.6 Graduation to Phase III

[At the time of preparing this version of the SAP, all agents that had been initiated in phase II evaluation have been evaluated for graduation to phase III evaluation or a decision has been made not to evaluate them for graduation. The following therefore provides a summary of the approach to graduation that is in protocol version 8.0 recognizing that there are currently no further agents being considered for graduation].

Each investigational agent that is being considered for evaluation in phase III will be evaluated for safety, for activity in reducing COVID-19 symptoms and hospitalization/death, and for activity in reducing SARS-CoV-2 RNA shedding. An analysis to determine if an agent should graduate from phase II, and enter phase III, will be conducted when 220 participants assigned to the agent or concurrent placebo in phase II evaluation have completed their Day 7 evaluations and have the required data available in the database. Additional interim graduations may be assessed for some agents, see agent-specific appendices in the protocol for details.

The DSMB will review unblinded data and make recommendations to NIAID (as trial sponsor) and to the TOC, indicating whether graduation criteria have been met. The recommendation for an agent to enter phase III evaluation will be made by the TOC in discussion with the collaborating company; the collaborating company that is responsible for the agent will decide whether or not to adopt the recommendation. The TOC and collaborating company will also consider which dose to recommend for evaluation in phase III, for investigational agents with more than one dose under evaluation in phase II. NIAID/DAIDS, as the sponsor of the study, will make the final determination regarding graduation of the study product.

The TOC may recommend an agent move directly into phase III, without evaluation in phase II in ACTIV-2, if there is sufficient safety and efficacy data supporting phase III evaluation available from outside of the trial. These agents will not undergo graduation analyses.

Graduation criteria and statistical considerations are discussed in the Graduation Rules SAP.

3 Outcome Measures

All outcome measures are copied from the protocol version 8.0. Only outcome measures addressed in this SAP are included below. See protocol section 10.2 for additional outcome measures.

3.1 Primary Outcome Measures: Phase III

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Efficacy: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo [active comparator intervention under protocol version 7.0]. [For Primary Objective 4]

   Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic.

3.2 Primary Outcome Measures: Phase II

1) Safety: New Grade 3 or higher AE through 28 days. [For Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Clinical (Symptom Duration): Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) based on self-assessment. [For Primary Objective 2]

   Duration defined as the number of days from start of investigational treatment to the first of two consecutive days when all symptoms scored as moderate or severe at study entry (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at study entry (pre-treatment) are scored as absent. The targeted symptoms are fever or feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each symptom is scored daily by the participant as absent (score 0), mild (1), moderate (2) and severe (3).

3) Virologic: At each of days 3, 7 and 14 quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs. [For Primary Objective 3]

3.3 Secondary Outcome Measures

Safety

1) Phase II only: New Grade 2 or higher AE through 28 days.
[Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

2) Phase II only: New Grade 2 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

3) Phase III only: New Grade 3 or higher AE through week 24.
[Supportive of Primary Objective 1, with follow-up beyond day 28]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

Clinical Symptoms

4) Phase III only: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2]

   Duration defined as the same as the primary phase II clinical (symptom duration) outcome.

5) Phase II and III: Duration of targeted COVID-19 associated symptoms from start of investigational agent (day 0) through day 28 based on self-assessment.
[Supportive of Primary Objective 2 and for Secondary Objective 8]

   Duration defined as the number of days from start of investigational treatment to the first of four consecutive days when all symptoms are scored as absent. Targeted symptoms as defined in the primary phase II clinical (symptom duration) outcome.

6) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28. [Supportive of Primary Objective 2]

Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of two consecutive days that a participant reported return to usual (pre-COVID) health.

7) Phase II and III: Time to self-reported return to usual (pre-COVID-19) health as recorded in a participant's study diary through day 28.
[Supportive of Primary Objective 2]

   Time to self-reported return to usual health defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

8) Phase II and III: COVID-19 severity ranking based on symptom severity scores over time during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention, hospitalization, and death. [For Secondary Objective 1].

   Participants who are alive at 28 days and not previously hospitalized, the severity ranking will be based on their area under the curve (AUC) of the daily total symptom score associated with COVID-19 disease over time (through 28 days counting day 0 as the first day) where the total symptom score on a given day is defined as the sum of scores for the targeted symptoms in the participant's study diary (each individual symptom is scored as absent (score 0), mild (1) moderate (2) and severe (3)). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

9) Phase II and III: Progression through day 28 of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at study entry, prior to start of investigational agent or comparator intervention. [For Secondary Objective 2]

10) Phase II and Phase III: Oxygen saturation (i.e., pulse oximeter measure) categorized as <96 versus ≥96% through day 28. [For Secondary Objective 6]

11) Phase II only: Level (quantitative) of oxygen saturation (i.e., pulse oximeter measure) through day 28. [Supportive of Secondary Objective 6]

<u>Virology</u>

12) Phase III (Active-Controlled [protocol version 7.0] and placebo-controlled [protocol version 8.0]) only: Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA from staff-collected NP swabs at day 3. [Support of Primary Objective 3]

13) Phase II and III: Level (quantitative) of SARS-CoV-2 RNA from staff-collected NP swabs at days 3, 7, and 14 in phase II and at day 3 in phase III.8.
[For Secondary Objective 3]

14) Phase II only: Area under the curve and above the assay lower limit of quantification of quantitative SARS-CoV-2 RNA over time from staff-collected NP swabs at days 0, 3, 7, and 14. [Supportive of both Primary Objective 3 and Secondary Objective 3]

Efficacy

15) Phase II only: Death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

    Hospitalization is defined as the same as the primary phase III outcome.

16) Phase II and III: Death due to any cause during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Primary Objective 4]

17) Phase II and III: Death due to any cause or hospitalization due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

18) Phase II and III: Death due to any cause or hospitalization due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 8, with follow-up beyond day 28]

    Hospitalization is defined as the same as the primary phase III outcome.

19) Phase II and III: Death due to any cause during the 24-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

20) Phase II and III: Death due to any cause during the 72-week period from and including the day of the first dose of investigational agent or comparator intervention. [Supportive of Secondary Objective 8, with follow-up beyond day 28]

21) Phase III: Death due to any cause or hospitalization due to any cause, excluding hospitalizations that are deemed unrelated to COVID-19, during the 28-day period from and including the day of the first dose of investigational agent or comparator intervention. [For Secondary Objective 9]

    Hospitalization is defined as ≥24 hours of acute care, in a hospital or similar acute care facility, including Emergency Rooms or temporary facilities instituted to address medical needs of those with severe COVID-19 during the COVID-19 pandemic. An Adjudication committee is evaluating the relatedness of hospitalization due to COVID-19.

3.4 Other Outcome Measures

1) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 28 days from start of investigational agent or comparator intervention. [For Exploratory Objective 1]

2) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 28 days from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1]

3) Phase II and III: New SARS-CoV-2 positivity among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [For Exploratory Objective 1, with follow-up beyond day 28]

4) Phase II and III: New SARS-CoV-2 positivity or COVID-19 symptoms among household contacts through to 24 weeks from start of investigational agent or comparator intervention. [Supportive of Exploratory Objective 1, with follow-up beyond day 28]

5) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through day 28. [For Exploratory Objective 4]

   Ordinal scale defined as:

   Death
   Hospitalized, on invasive mechanical ventilation or ECMO;
   Hospitalized, on non-invasive ventilation or high flow oxygen devices;
   Hospitalized, requiring supplemental oxygen;
   Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

6) Phase II and III: Duration of hospital stay among participants who become hospitalized through day 28. [For Exploratory Objective 4]

7) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through day 28. [For Exploratory Objective 4]

8) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through day 28. [For Exploratory Objective 4]

9) Phase II and III: Worst clinical status assessed using ordinal scale among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

Ordinal scale defined as:

Death
Hospitalized, on invasive mechanical ventilation or ECMO;
Hospitalized, on non-invasive ventilation or high flow oxygen devices;
Hospitalized, requiring supplemental oxygen;
Hospitalized, not requiring supplemental oxygen (COVID-19 related or otherwise).

10) Phase II and III: Duration of hospital stay among participants who become hospitalized through week 24.
[For Exploratory Objective 4, with follow-up beyond day 28]

11) Phase II and III: ICU admission (yes versus no) among participants who become hospitalized through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

12) Phase II and III: Duration of ICU admission among participants who are admitted to the ICU through week 24. [For Exploratory Objective 4, with follow-up beyond day 28]

4 Statistical Principles

4.1 General Considerations

The following analysis populations are defined for a given investigational agent:

- Screened Population: All participants who were screened for enrollment into the study, between the time of screening of the first and last participants who were eligible to be randomized to the given Investigational Agent Group.

- Randomized Population: All participants who were enrolled and were eligible to be randomized to the given Investigational Agent Group, and were actually randomized either to the investigational agent or to its comparator intervention (placebo or active comparator, as appropriate for the agent and phase of evaluation).

- Treated Population: All participants in the Randomized Population who received any investigational agent or its comparator agent (this is a modified intent-to-treat [mITT] population).

In general, the Treated Population is the focus of randomized comparisons to evaluate the safety and efficacy outcomes of an investigational agent versus its comparator intervention. In all analyses of a given investigational agent, the comparison group will include all participants who were concurrently randomized to the comparator intervention, who were also eligible to have received the investigational agent of interest. For the placebo-controlled trials, the comparison group will pool across all relevant placebos (i.e. including the placebo for the agent of interest and the placebos for other agents). For the primary placebo-controlled analysis of a specific investigational agent, a supplemental analysis may be undertaken that restricts the comparison group to include only participants who received the placebo for that specific investigational agent.

Study visit windows for reporting are based on the Schedule of Evaluations (SOE) defined in the protocol (in person visits shown in the below table) and will be derived based on the evaluation/specimen date and study treatment initiation date (at interim analyses, if not available, study start date will be used). In the event that multiple results fall within the same analysis window, the one closest to the target time point will be prioritized, or if equidistant from the target time point, the earlier result will be prioritized. For interim analyses, if a result does not fall in an analysis window, the visit label will be used to identify the target time point.

SOE Visit	Protocol Range (Days)	Analysis Range (Days)	Analysis Window (Days)
Screening	-2, 0	-10, 0	-10, 0
Day 0*	0	-1, 0	-1, 0
Day 3	2, 4	1, 4	-2, +1
Day 7	5, 9	5, 10	-2, +3
Day 14	12, 16	11, 21	-3, +7
Day 28	28, 32	22, 38	-6, +10
Week 12	77, 91	56, 112	+/- 28
Week 24	161, 175	140, 196	+/- 28
Week 36	245, 266	224, 280	+/- 28
Week 48	329, 350	308. 364	+/- 28
Week 72	497, 518	476, 532	+/- 28

*The Day 0 analysis window is designed to capture data in scenarios where randomization occurs on the day prior to treatment initiation. Evaluations that occur on Day 0, post-treatment initiation (e.g., vital signs evaluations), will consider the time of the evaluation compared to the time of treatment administration (and will be presented as 'Day 0' with the relative time). Windows cited above do not apply to data with daily collections (i.e., diary cards or nasal swabs).

Key study visits are Entry (Day 0), day 28, week 24:

Entry (Day 0): First dose of investigational agent/comparator intervention occurs.

Baseline is defined as the last available measure prior to the initiation of investigational agent/placebo.

Day X: Last day of investigational agent/comparator intervention.

Value of X depends on agent: see protocol appendices for details for each specific investigational agents.

Day 28: Last day primary outcome may occur.

Week 24: Key visit for evaluating longer-term outcomes for all agents (note: some agents may have follow-up beyond week 24).

Week 72: Key visit for evaluation longer-term efficacy and safety for some agents (see agent specific appendices).

Statistical comparison across randomized arms of baseline characteristics are not planned because the study is randomized; hence, any differences should reflect chance variation. In addition, comparisons between investigational agents are not planned. Control of the Type I error rate will be undertaken separately for each investigational agent, and not across all investigational agents (i.e., not for the experiment-wise or family-wise error rate of the study).

Analyses of primary and secondary outcomes will not adjust for multiple comparisons. Analyses of primary outcomes will adjust for the multiple interim reviews using group sequential methods.

Continuous variables will be summarized using mean, standard deviation, median, interquartile range (Q1 and Q3), 10^{th} and 90^{th} percentile, and min and max; categorical variables will be summarized using frequency and percentage.

NIH requires that the primary outcomes also be summarized by randomized arm by sex/gender and by race/ethnicity, and that treatment interactions with sex/gender and race/ethnicity be evaluated.

SARS-CoV-2 RNA results may be below the assay lower limit of quantification (LLoQ) or above the upper limit of quantification (ULoQ). Values below the LLoQ or above the ULoQ will generally be considered as censored observations in statistical analyses (with left censoring at the LLoQ and right censoring at the ULoQ, respectively). However, if necessary for any analyses (and for graphical presentations), values may be imputed in the following manner:

- Values below the LLoQ, but above the limit of detection (LoD) will be imputed as half the distance from the log-10 transformed LoD to the log-10 transformed LLoQ
- Values below the LLoQ and below the LoD will be imputed as half the distance from zero to the log-10 transformed LoD;
- Values above the ULoQ will be imputed as one unit higher than the log-10 transformed ULoQ; actual values obtained from assay reruns with dilution will be used instead, if available.

Virology results generated from specimens with the following conditions reported in the database will be excluded from analyses:

- Thawed;
- Invalid Specimen;
- Quantity Not Sufficient;
- Destroyed.

Note: Samples with the condition code ‘NOT’ were also to be excluded per the trial sponsor but this code indicates that the specimen was not tested. Thus, no result is expected and no exclusion is needed.

5 Analysis Approaches

All analyses addressing the primary and secondary objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach, i.e. using the Treated Population. Note that according to the protocol, participants who are randomized but do not start investigational agent or comparator intervention are not followed.

Participants who have protocol violations, such as those who start investigational agent or comparator intervention outside of the protocol-defined study windows, or who are found to be ineligible, will be included in the analysis on the basis that they were considered part of the target population at the time of randomization and start of treatment.

For agents in phase II evaluation, participants who were at “higher” risk of progression to severe COVID-19 when eligible and enrolled under an earlier version of the protocol will be included in all analyses. Similarly, participants who were eligible and enrolled with longer than 7 days from symptom onset to study entry will be included in all analyses.

5.1 Analyses of the Primary Objectives

5.1.2 Phase III Primary Objective for Efficacy: Placebo-Controlled Superiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the placebo-controlled superiority design. Further details are provided after the table.

<table>
<tr><td colspan="2">Phase III Primary Objective for Efficacy—Placebo-Controlled Superiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.</td></tr>
<tr><td>Estimand description</td><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.</td></tr>
<tr><td>Treatment</td><td>Investigational agent or placebo.</td></tr>
<tr><td>Target population</td><td>Analysis set (analysis population)</td></tr>
<tr><td>Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 10** days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry</td><td>Treated Population</td></tr>
<tr><td>Variable(s)</td><td>Outcome measure(s)</td></tr>
<tr><td>Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo (coded as 1 if participant died or was hospitalized, and 0 otherwise).

To handle censoring due to loss to follow-up before 28 days in statistical analysis, a time variable for study day of hospitalization/ death or censoring (earlier of 28 days or day of last contact with participant) is also needed.</td><td>Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or placebo.</td></tr>
<tr><td>Handling of intercurrent events</td><td>Handling of missing data</td></tr>
<tr><td>None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of an agent/placebo).</td><td>Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as (non-informatively) censored at the date last known to be alive.</td></tr>
<tr><td>Population-level summary measure</td><td>Analysis approach</td></tr>
<tr><td>Ratio (for investigational agent divided by placebo group) of cumulative probability of death or hospitalization over 28 days.</td><td>Ratio (for investigational agent divided by placebo group) of the cumulative proportion dying or being hospitalized at day 28 obtained using Kaplan-Meier estimation using the indicator variable for hospitalization/death and the time variable described above. See text for further details.</td></tr>
<tr><td colspan="2">* * This was changed from 10 days under protocol version 2 and protocol version 3, to 8 days under LOA#1 to protocol version 3, (also applies to protocol version 4 and 5), and to 7 days under protocol version 6 and subsequent protocol versions.</td></tr>
</table>

Analysis Approach

The analysis of the primary efficacy outcome in phase III will compare the cumulative proportion of participants hospitalized or died (due to any cause), from day 0 through day 28, between randomized arms using a ratio of proportions; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included. The cumulative proportion will be estimated for each randomized arm using Kaplan-Meier methods to account for losses to follow up (and differential follow-up at the interim reviews). For analysis purposes, the integer scale will be used as the time scale, where study day 0 is the day of start of investigational agent or placebo, study day 1 is considered day 1, and study day 28 is considered day 28; if an event occurs on day 0 then event time will be set to 0.5 for analysis. Participants will have follow-up censored at the date they were last known to be alive and not hospitalized through day 28. The primary analysis assumes non-informative censoring.

The absolute difference in the estimated log-cumulative proportion will be calculated between randomized arms; a 95% CI will be obtained for this difference in log-cumulative proportion calculated using a variance for this difference being the sum of the variances for each randomized arm obtained using Greenwood's formula. Results will be anti-logged to give the estimated ratio of cumulative proportions through day 28 (investigational agent vs placebo) and associated 95% CI. Two-sided 95% confidence intervals (CIs) and p-value (for the test of no difference between groups) will be obtained, which adjust for the interim analyses; a nominal 95% CI and p-value will also be provided.

It is possible, particularly at interim analyses for DSMB reviews, that the number of hospitalizations/deaths in an arm (investigational agent or placebo) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events (hospitalizations/deaths) in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using Greenwood's formula to calculate confidence intervals for the difference between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the primary comparisons. The third sensitivity analysis is an exploratory analysis.

1) Evaluate the composite outcome of being hospitalized, dead, or loss-to-follow-up.

   Approach: Repeat the primary analysis, but assume all participants who prematurely discontinued study follow-up prior to day 28 and who were unable to be contacted by the site to ascertain outcomes after discontinuation, had a primary event at day 28. See sensitivity analysis number 3 below for evaluating the potential impact of differential loss to follow-up.

2) Evaluate the impact of participants enrolling from the same household.

   Approach: Repeat the primary analysis only including the first participant who enrolled from each household.

   In the event that interpretation of results for the primary analysis differs substantially from the results from this sensitivity analysis, analysis methods that account for clustering will be considered, if feasible.

3) Exploratory: Evaluate the impact of differential loss-to-follow-up (LTFU).

   Approach: In the event that interpretation of the results for the primary analysis differs substantially between the primary analysis and the first sensitivity analysis, the impact of participants being LTFU will be explored using IPCW potentially using both pre-treatment variables and variables after starting study treatment to determine weights. The primary analysis will be repeated but, within each group, participants who are not LTFU will be weighted using IPCW determined by baseline variables that predict LTFU.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 17, 18, 19, 20 and 21, evaluate the proportion of participants who are hospitalized or died through week 24, the proportion who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In these analyses, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days) or through 504 days (i.e. 72 times 7 days).

Secondary outcome 15 is included to assess the phase III primary efficacy outcome of hospitalization or death during phase II. This outcome will be analyzed in the same manner as the primary efficacy outcome in phase III if there are 5 or more participants who died or were hospitalized in each arm. If not, the number of deaths and hospitalizations will be summarized and compared between arms using Fisher’s exact test.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented for each subgroup. Within each subgroup, the difference between randomized arms in the log-proportion will be estimated, and compared between subgroups by constructing a test of interaction and 95% confidence interval. This will be implemented by determining the difference between subgroups of the differences between randomized arms, and the variance of the difference will be determined by summing the variance of the subgroup-specific variances. In the event that the number of events in a subgroup in either the investigational arm or placebo arm is low (less than 5), descriptive summaries of the number of hospitalizations and deaths by subgroup and arm will be provided. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/placebo Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Phase III Primary Objective for Efficacy: Active-Controlled Non-Inferiority Evaluation

The following table summarizes the primary efficacy objective in phase III and the associated estimand under the active-controlled non-inferiority design. Further details are provided after the table.

Phase III Primary Objective for Efficacy—Active-Controlled Non-Inferiority Evaluation: To determine if the investigational agent will prevent the composite endpoint of either hospitalization due to any cause or death due to any cause through study day 28.	
Estimand description	Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization through day 28, among adults with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry.
Treatment	Investigational agent or active comparator agent (casirivimab and imdevimab).

Target population	Analysis set (analysis population)
Adults (≥ 18 years of age) with documented positive SARS-CoV-2 molecular test results collected within 240 hours (10 days) prior to study entry with no more than 7 days of symptoms of COVID-19 prior to study entry, and with presence of select symptoms within 24 hours of study entry	Treated Population
Variable(s)	**Outcome measure(s)**
Indicator variable for death due to any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent (coded as 1 if participant died or was hospitalized, and 0 otherwise).	Death due any cause or hospitalization due to any cause during the 28-day period from and including the day of the first dose of investigational agent or active comparator agent.
Handling of intercurrent events	**Handling of missing data**
None. A treatment policy strategy is being taken to evaluate treatment effects irrespective of intercurrent events (e.g. irrespective of whether a participant received the complete dose(s) of the agent to which they were randomized.	Participants who discontinued follow-up before day 28 without previously dying or being hospitalized will be considered as not having an event after the date last known to be alive.
Population-level summary measure	**Analysis approach**
Difference (for investigational agent minus active comparator agent) of probability of death or hospitalization over 28 days.	Difference (for investigational agent minus active comparator agent) of the proportion dying or being hospitalized at day 28. See text for further details.

Analysis Approach

The analysis of the primary efficacy outcome in phase III will evaluate the absolute difference in proportion of participants hospitalized (due to any cause) or died (due to any cause), from day 0 through day 28, between randomized arms; hospitalizations that begin on day 28 and deaths that occur on day 28 will be included.

Inference will be based on constructing a two-sided exact 95% confidence interval for the absolute difference in proportions (proportion for the investigational agent minus the proportion for the active comparator agent). If this confidence interval is entirely below the non-inferiority margin of 3%, then a conclusion of non-inferiority of the investigational agent compared with the active

comparator agent will provide reasonable evidence that the investigational agent is effective against COVID-19.

The exact 95% confidence interval will be calculated using the method of Chan and Zhang [Biometrics 1999;55:1201-09] as implemented, for example, in StatXact PROC BINOMIAL for SAS [StatXact 12 PROCs for SAS Users Manual. Cytel Inc., Cambridge, MA; 2019]. This method inverts two one-sided hypothesis tests (with one-sided error rate of 0.025 each) to obtain the confidence interval so providing a confidence interval-based method which preserves the type I error rate in establishing non-inferiority to be 0.025. To preserve confidence interval coverage (and type I error rate for assessing non-inferiority) over multiple interim analyses, the confidence interval will be calculated using a “repeated” confidence interval approach with spending of error rate at each interim analysis using the Land and DeMets approach with an O’Brien and Fleming spending function.

In essence, basing the comparison of treatment groups on the simple proportion of participants who were hospitalized or died assumes that participants who are lost to follow-up before 28 days without prior hospitalization were not hospitalized and did not die by 28 days. The decision to use the simple proportion for analysis rather than use, for example, a Kaplan-Meier estimate of the cumulative proportion of participants hospitalized or dying during the first 28 days of follow-up to account for losses to follow-up was taken for multiple reasons. First, in ACTIV-2 and other COVID-19 trials, most hospitalizations and deaths occur during the first two weeks of follow-up and the study has been designed to have regular contact with participants or their secondary contacts so as to maximize ascertainment of hospitalization and death information. Second, loss to follow-up has been low in the ACTIV-2 study: approximately 3% among higher risk participants. Third, with the very low rates of hospitalization/death expected (e.g., 2.3% for the active comparator agent), confidence interval coverage (and type I error rates) are better preserved at their desired levels through the use of exact statistical methods for analyzing proportions than is achieved using asymptotic statistical methods based on Wald-type analyses using Greenwood’s formula to obtain standard errors for Kaplan-Meier estimates. To assess the potential impact of loss to follow-up (assumed to be non-informative) on the interpretation of results, the following sensitivity analyses will be undertaken, repeating the primary analysis repeated with:

(a) a comparison of the simple proportions using a Wald-based confidence interval; and

(b) a comparison of proportions estimated using Kaplan-Meier methods (with censoring of follow-up at the earlier of day 28 and the time that a participant was last known to be alive) using a Wald-based confidence interval with standard error based on Greenwood’s formula.

Supportive Analyses

Secondary Outcome 16 is included as supportive to the primary efficacy outcome. The cumulative proportion of participants dead (due to any cause) by day 28 will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 16, 17, 18, 19, 20 and 21 evaluate the proportion of participants who die through to day 28, the proportion who are hospitalized or died through week 24, the proportion

who are hospitalized or died through week 72, the proportion who died (due to any cause) through week 24, the proportion who died (due to any cause) through week 72, and the proportion who died or were hospitalized excluding hospitalizations deemed unrelated to COVID-19 though day 28. These outcomes will be analyzed in the same manner as the primary efficacy outcome. In the sensitivity analyses based on Kaplan-Meier estimates, however, participants will have their follow-up censored at the date they were last known to be alive and not hospitalized (or date they were last known to be alive) through 168 days (i.e. 24 times 7 days for outcomes through to 24 weeks) or through 504 days (i.e. 72 times 7 days for outcomes through to 72 weeks).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary outcome will be assessed among different subgroups. The same approach outlined for the primary analysis will be implemented for each subgroup. However, these analyses are likely to involve small numbers of events in most or all subgroups and hence have very limited precision. Because of this, any assessment of treatment by subgroup interaction, if undertaken, will be considered exploratory. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/active comparator Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.2 Primary Safety (Phase II and III)

Analysis Approaches

Occurrence of any new Grade 3 or higher AE through 28 days will be analyzed in the following manner. The proportion of participants who experienced a new Grade 3 or higher AE will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. A 95% confidence interval for the risk ratio and a two-sided p-value from a Wald test of the null hypothesis that the risk ratio is one will also be provided. In the event the log-binomial regression model fails to converge or has questionable convergence, a Poisson regression model with robust variance and log-link will be used instead.

In addition, the absolute difference in proportion of participants who experienced a new Grade 3 or higher AE (or new Grade 2 or higher AE) will be calculated, with associated 95% confidence interval (calculated using the normal approximation to the binomial distribution).

It is possible, particularly at interim analyses for DSMB reviews, that the number of Grade 3 or higher AEs in an arm (investigational agent or comparator intervention) will be very small and hence the asymptotic (large sample size) statistical theory underpinning the above statistical analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare proportion between arms will be adopted instead of using the log-binomial regression model and normal approximation to the binomial distribution to calculate confidence intervals for the relative and absolute differences between arms and associated p-values. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

In placebo-controlled evaluations, because some agents may be administered using injections or infusions and others will not be, the primary safety analysis may be repeated on the subset of the Treated Population that received the investigational agent of interest or the placebo for that specific agent.

Supportive Analyses

Secondary Outcome 1 is included as supportive to the primary safety outcome in phase II. This outcome evaluates the occurrence of new Grade 2 or higher AEs through 28 days, and will be analyzed in the same manner as the primary outcome.

Secondary Outcomes 2 and 3, which are included in support of the primary safety objective, evaluate the occurrence of new Grade 2 or higher AEs (in phase II) and Grade 3 or higher AEs (in phase III) through week 24. These outcomes will be analyzed in the same manner as the primary safety outcomes.

Additional longer-term safety outcomes may be assessed, see agent-specific appendices for details.

5.1.3 Primary Clinical Symptoms (Phase II)

Analysis Approaches

The targeted symptoms considered in evaluating the primary symptom outcome are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent, mild, moderate or severe from day 0 (pre-treatment) through day 28.

The primary symptom outcome measure is the time to when all targeted symptoms are sufficiently improved or resolved for two consecutive days from their status at day 0 (pre-treatment). Specifically, it is defined as the number of days from start of investigational agent (day 0, pre-treatment) to the first of two consecutive days when all symptoms scored as moderate

or severe at day 0 (pre-treatment) are scored as mild or absent, AND all symptoms scored as mild or absent at day 0 (pre-treatment) are scored as absent.

Statistically, this is a time-to-event (TTE) variable, potentially involving censoring due to loss-to-follow-up or if a participant did not meet the outcome criteria for symptoms sufficiently improved/resolved during the 28 days of completing the diary. Censoring of follow-up for the TTE outcome measure will occur on the last day that the TTE outcome measure could have been achieved. Specifically, as two consecutive days of symptoms meeting the outcome measure criteria are required, censoring would be on the day before the last day of completion of the diary card (e.g., this would be day 27 for participants with complete diaries through day 28, as meeting the criteria requires completion of the diary on both day 27 and day 28). Descriptive analyses for this TTE outcome measure will be undertaken using Kaplan-Meier methods including "survival" functions and/or cumulative incidence plots, and associated summary statistics (median [quartiles] with 95% confidence interval; and estimated % not meeting outcome measure criteria by 28 days with a 95% confidence interval). Comparison of the distribution of the TTE outcome measure between investigational agent and comparator intervention arms will be undertaken using Wilcoxon's test adapted for handling censored data (the Gehan-Wilcoxon test) using a two-sided Type-I error rate of 5%.

For each participant, the symptom data that contribute to the calculation of the TTE outcome measure and the censoring time (and associated censoring indicator variable) can be described as a panel of evaluations (absent/mild/moderate/severe) for each of 13 targeted symptoms on each of 29 days (day 0 through day 28). The following general principles will be applied for the handling of deaths, hospitalizations, and missing data:

- **Deaths**. Participants who die without previously achieving the TTE outcome (i.e. without two consecutive days of symptoms improved/resolved), will be retained in the risk set for the TTE outcome, but without achieving the TTE outcome, from the day of death (or the day after death if the diary was completed on the day of death) through to and including study day 27. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days, over time.

- **Hospitalizations.** Participants who are hospitalized without previously achieving the TTE outcome measure will be retained in the risk set for the TTE outcome, but without having the TTE outcome, from all days hospitalized (including day of admission if no diary was completed that day, and including day of discharge if no diary was completed that day). As the protocol does not expect that diaries are completed during hospitalization, diary evaluations that are completed from the day after admission to the day before discharge will be ignored. The underlying premise is that participants have not achieved symptom improvement/resolution while hospitalized.

- **Losses to Follow-up and Early Termination of Evaluation of Targeted Symptoms.** Participants who are lost to follow-up or who terminate providing evaluations of the targeted symptoms in their study diaries before day 28 for any reason have monotonic

missing data (i.e. a sequence of missing values during follow-up through to and including day 28). For these participants, the TTE outcome measure will be censored at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for one or more targeted symptoms). For the special case of participants who have no evaluations of targeted symptoms in their study diaries from the day of hospital discharge through to day 28 for any reasons, the TTE outcome measure will be censored at the day before discharge. If the participant withdraws from the study while hospitalized and therefore no date of discharge is available, then the TTE outcome measure will be censored on the day before withdrawal from the study. These criteria for censoring assume that the censoring is non-informative about when the TTE outcome would have been met if diaries had been fully completed after the last diary entry for one or more targeted symptoms (or after hospitalization or after withdrawal from the study during hospitalization).

- **Intermittent Missingness**. Participants who have intermittent missing evaluations for a specific symptom (i.e. one or more successive evaluations with preceding and succeeding evaluations for the same symptom) will have the missing evaluation(s) imputed as the worst of the preceding and succeeding evaluations for the same symptom. There may be no impact of this on the TTE outcome if evaluations of other symptoms are completed and do not meet the TTE outcome during the period of missingness for the specific symptom. If there is an impact, it may be to move the TTE outcome earlier (than if the evaluations had been done) if both the preceding and succeeding evaluations for the specific symptom meet the criteria for improvement/ resolution; and, conversely, to move the TTE outcome later (than if the evaluations had been done), if both the preceding and succeeding evaluations for the specific symptom don't meet the criteria for improvement/resolution.

- **Missing Day 0 Evaluation.** If the evaluation at day 0 is missing for a given symptom and there is at least one evaluation provided for that same symptom during follow-up, then the missing evaluations at day 0 and subsequently through to the first evaluation will be imputed as "mild". The choice of imputation as "mild" is based on the fact that among early participants in ACTIV-2, the median evaluation given to any specific symptom at day 0 was "mild". This imputation means that the improvement/resolution criteria cannot be met based on these imputed data (as the criteria for a mild symptom at day requires resolution to absent). The impact of this may be to move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "absent" or "mild"; and it may also move the TTE outcome later (than if the evaluations had been done) if the true day 0 evaluation would have been "moderate" or "severe" as the imputed "mild" symptom at day 0 must resolve to absent whereas a true "moderate" or "severe" symptom only need to resolve to "mild".

Appendix 1 includes a detailed description of an algorithm for handling missing data following these general principles that can be implemented programmatically.

Supportive Analysis

The analysis will be repeated using the same approach as described above (including handling of deaths, hospitalizations and missing data) for a similar TTE outcome measure defined as time to (a) two consecutive days with resolution of all targeted symptoms to “absent”, and (b) four consecutive days with resolution of all targeted symptoms to “absent” (i.e., secondary outcome measure 5). For these two outcomes, as for the primary symptom outcome measure, the first day that a participant may meet this outcome will be day 1 (i.e. if all targeted symptoms are “absent” on (a) both day 1 and day 2, or (b) on days 1, 2, 3 and 4).

Sensitivity Analysis

It is possible that a participant may meet the primary TTE symptom outcome measure and subsequently be hospitalized or die. To assess how sensitive the primary symptom outcome results might be to this form of improvement and then deterioration, the primary analysis maybe repeated with participants who are hospitalized or who die by day 28 kept in the risk set through to day 28 without meeting the improvement/resolution outcome (i.e. assuming that they did not achieve this outcome if they actually did). It is recognized that this adaptation means that the outcome measure being analyzed is not a true TTE outcome measure but this analysis does allow an assessment of the sensitivity of results to the handling of participants who are hospitalized or who die. [Note: this sensitivity analysis was suggested by the Food and Drug Administration].

No additional sensitivity analyses are currently specified for this outcome measure. In part, this is because the proportion of participants enrolled early in ACTIV-2 who were lost to follow-up or who had extensive missing diary evaluations has been very low, and not all loss to follow-up or missingness patterns affect the determination of the TTE outcome. If necessary, exploratory sensitivity analyses will be undertaken to explore sensitivity of interpretation of results for the comparison of investigational agent to comparator intervention to losses to follow-up and/or missing data but these may need specification based on the form of missingness identified.

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary symptom outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) ‘Risk of Severe Disease’ Stratification [this may not be pursued for agents which predominantly enrolled participants who were at ‘lower’ risk for severe COVID progression]
5) Age Group (<60, ≥60)

6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing).

5.1.4 Primary Virologic (Phase II)

Analysis Methods

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ in NP swabs at each scheduled measurement time (entry and days 3, 7, and 14).

The proportion of participants with SARS-CoV-2 RNA < LLoQ will be compared between randomized arms using log-binomial regression for repeated binary measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI and two-sided p-value) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ. However, in the event that there is a non-negligible proportion of baseline results <LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as "0" if above LLoQ, and "1" if below LLoQ).

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in the primary analysis. Sensitivity analyses will address possible informative missingness (see below).

If there is a need to conduct analyses of interim data (e.g. if requested by the DSMB), then the primary statistical analysis described above may be sensitive to small numbers of participants with data available at some measurement times. Because of this, such interim analyses will be undertaken using log-binomial models fit separately at each time point. If at a given time point, the number of participants with SARS-CoV-2 RNA < LLoQ or, conversely the number with SARS-CoV-2 RNA ≥ LLoQ in an arm (investigational agent or comparator intervention) is small, the asymptotic (large sample size) statistical theory underpinning these model-based analyses may be questionable. To address this, using a standard rule of thumb, if there are fewer than 5 events in either arm, inference based on Fisher's exact test to compare arms will be adopted instead of using the log-binomial regression model. If there are no participants have SARS-CoV-2 RNA <LLoQ (or all participants have SARS-CoV-2 RNA ≥ LLoQ) in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.
2) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used (as implemented in SAS PROC GEE [Lin G, Rodriguez RN. Weighted methods for analyzing missing data with the GEE procedure. Paper SAS166-2015. 2015.]; based on Robins and Rotnitzky. Journal of the American Statistical Association. 1995 Mar 1;90(429):122-9; Preisser, Lohman, and Rathouz. Statistics in Medicine. 2002 Oct 30;21(20):3035-54).
3) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The primary analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

Subgroup Analyses

To evaluate the effect of the investigational agent in specific populations, the primary virology outcome will be assessed among different subgroups. The same approaches outlined for the primary analysis will be implemented within each subgroup; formal comparisons across subgroups will not be done in phase II analyses. Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) 'Risk of Severe Disease' Stratification
   [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
5) Age Group (<60, ≥60)
6) Co-morbidity Status (no comorbidities, at least one comorbidity) [this may not be pursued for agents which predominantly enrolled participants who were at 'lower' risk for severe COVID progression]
7) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
8) Country (U.S., non-U.S.)
9) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2 Analyses of Secondary Objectives

Analysis Population

The analyses of the COVID-19 symptoms will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed.

5.2.1 Secondary Clinical Symptoms

Analyses Methods

Duration of Clinical Symptoms

Duration of clinical symptoms in phase III will be analyzed in the same manner as the primary phase II clinical symptom outcome.

Progression of Symptoms

Progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary on day 0 (pre-treatment) on or before day 28 (i.e., absent to at least mild, mild to at least moderate, or moderate to severe) will be analyzed in the following manner. The proportion of participants who progressed will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Participants who do not report worsened symptoms in study diaries, but are hospitalized or die in the first 28 days will be counted as having progression of symptoms in this analysis. Missing symptom scores not due to hospitalization or death will be imputed in the same manner as the primary symptom duration outcome (see above).

Return to Usual Health

The study diary includes a question: "*Have you returned to your usual (pre-COVID) health today?*" which is answered each day with possible responses "*yes*" or "*no*". Duration of time without self-reported return to usual health is defined as the number of days from start of treatment to the first of two consecutive days that self-reported return to usual health was indicated as "*yes*".

Analysis (including handling of hospitalizations, deaths and missing data) will follow the same approach as for the primary clinical symptom duration outcome measure as described above.

COVID-19 Severity Ranking

COVID-19 severity ranking will be summarized with descriptive statistics. Participant specific scores will be compared between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

The symptoms considered in calculating symptom duration are: feeling feverish, cough, shortness of breath or difficulty breathing at rest or with activity, sore throat, body pain or muscle pain/aches, fatigue (low energy), headache, chills, nasal obstruction or congestion (stuffy nose), nasal discharge (runny nose), nausea, vomiting, and diarrhea. Each of these symptoms is scored daily in a study diary by the participant as absent (score 0), mild (1) moderate (2) and severe (3) from day 0 (pre-treatment) to day 28.

COVID-19 severity ranking is defined as the participant-specific AUC of the total symptom score associated with COVID-19 disease, over time (through 28 days counting day 0 as the first day). For participants who are alive and were never hospitalized on or before day 28, the total symptom score on a particular day is the sum of scores for the targeted symptoms in the participant's study diary for that day. The AUC will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 28. The AUCs will be rescaled by time by dividing by 28, corresponding to the number of trapezoids created from daily diary cards between day 0 and day 28, in order to provide results on a symptom scale from 0 to 39.

Special considerations are made for participants who are hospitalized or die on or before day 28. Participants who are hospitalized or who die during follow-up through day 28 will be ranked as worse (i.e., worse severity) than those alive and never hospitalized through day 28 as follows (in worsening rank order): alive and not hospitalized at day 28; alive but hospitalized at day 28; and died on or before day 28. Programmatically, participants who were hospitalized, but are alive and no longer hospitalized at day 28 will be assigned an AUC (severity score) of 40, participants who are alive but remain hospitalized at day 28 will be assigned an AUC (severity score) of 41, and participants who die (regardless of when the death occurred through day 28) will be assigned a severity score of 42.

Participants who have incomplete diary cards for reasons other than hospitalization or death, and who are not subsequently hospitalized and do not die through day 28, will be addressed in the following manner:

1) Participants who are missing day 0 total symptom scores (i.e., participants who failed to complete the diary card on Day 0 and have no scores for any symptoms) will have their total symptom score imputed as the mean day 0 total symptom score among participants who report a total symptom score on day 0;
2) Participants who have some symptom scores missing at Day 0 (i.e., completed the diary card but did not score all symptoms) will have their total symptom score calculated as the mean of the available symptoms scores at Day 0, multiplied by 13;
3) Participants who stop completing their symptom diaries before day 28 will have their last total symptom score carried forward through day 28, and their AUC calculation done as noted above;
4) Participants who have diary cards with some, but not all symptom scores reported, their missing symptoms scores will be linearly interpolated based on the preceding and succeeding available scores for a given symptom, and their AUC calculation done as noted above;
5) Participants who have intermittent days with no symptom scores reported (i.e., all scores missing), their missing scores will be ignored in the AUC calculation, which is analogous to interpolating the total symptom scores.

Methods such as multiple imputation or IPCW may be considered if more than 10% of participants in either group stop completing their diaries before day 28 for reasons other than death or hospitalization.

To programmatically implement the imputation of the missing diary cards in order to calculate the AUC for participants who are not hospitalized and do not die by day 28, the following steps will be followed. First, imputation of total symptom scores will be done according to (1), (2), and (3). Next, (4) intermittent missing symptom scores for particular symptoms will be imputed using linear interpolation (see below formula) of the preceding and succeeding scores. Note: no imputation done for (5).

X = (Succeeding Score – Preceding Score) ÷ (Succeeding Day – Preceding Day)

Score on 1st Day missing = 1*X + Preceding Score

Score on 2nd Day missing = 2*X + Preceding Score

…..

Score on Zth Day missing = Z*X + Preceding Score.

Oxygen Saturation

Participants who are on supplemental oxygen at day 0 (pre-treatment) will not be included in these analyses.

Oxygen saturation will be analyzed in the same manner as the virology outcomes.

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with oxygen saturation ≥ 96% at each scheduled measurement time (day 0 [pre-treatment] and days 3, 7, 14, and 28).

The proportion of participants with any oxygen saturation values ≥ 96% will be compared between randomized arms using log-binominal regression for binary repeated measurements with log-link. This model will be fitted using generalized estimating equations (GEE) to handle the repeated measurements with an independence working correlation structure and robust standard errors. For each time point after starting treatment, the model will include a main effect for time (indicator variable for each evaluation time), and an interaction between time and randomized arm to evaluate differences between arms, and will adjust for baseline oxygen saturation level. The estimated adjusted relative risk of having oxygen saturation values ≥ 96% (and associated 95% CI) will be obtained for each measurement time from the model by taking the exponential of the time*randomized arm interaction parameter estimate (and associated 95% CI) for that measurement time. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

A joint test of randomized arm across the time points will also be assessed, with degrees of freedom determined by the number of time points included. With this model, the comparison between randomized arms will use a two-sided Wald-test with 5% type I error rate. Time points with zero events in either arm will not be included in the model (as estimation for such a model

may be problematic; however data for these time points will be included in a descriptive summary of results over time points).

Missing data are assumed to be missing completely at random (MCAR) and will be ignored in this analysis. Sensitivity analyses will address possible informative missingness (see below).

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare oxygen saturation levels (continuous) between randomized arms, separately at each post-entry study day.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the clinical symptoms outcomes.

Oxygen Saturation ≥ 96%

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96%, otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due to hospitalization or death, participants with missing oxygen saturation results will have their values imputed as <96%.
    - For non-monotonic missingness, participants with missing oxygen saturation results will have their values imputed as ≥96% if the preceding and succeeding results are ≥96% , otherwise the results will be imputed as <96%.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analyses

Duration of Symptoms

In support of the symptom duration outcome in phase III, the analysis will be repeated using the same approach described in the primary symptom duration analysis for a similar TTE outcome measure defined as time to two consecutive days with resolution of all targeted symptoms to “absent.” To address secondary objective 8, and in supportive of the symptom duration outcome in phase III, a similar TTE outcome measure will also be examined defined as time to four consecutive days with resolution of all targeted symptoms to “absent,” (i.e. secondary outcome measure 5).

Return to Usual Health

The analysis of return to usual health will be repeated using the same approach described above for a similar TTE outcome measures defined as the number of days from start of investigational treatment until the first of four consecutive days that a participant reported return to usual (pre-COVID) health.

COVID-19 Severity Ranking

To evaluate the effect of the investigational agent on COVID-19 symptom severity over different time-periods, analyses of COVID-19 severity ranking based on partial AUCs will also be examined. The time-periods considered include day 0 to day 7, day 0 to day 14, and day 0 to day 21. These analyses will compare participant specific AUCs between randomized arms using a two-sided Wilcoxon test with a 5% type I error rate.

For each time period, for participants who are alive and were never hospitalized in that time period (i.e., as of 7 days, 14 days, and 21 days), the severity ranking will be based on their AUC of the symptom score associated with COVID-19 disease over time (through day 7, 14, 21, respectively, counting day 0 as the first day) assigned as the sum of scores for the targeted symptoms in the participant's study diary. The AUCs will be calculated using the trapezoidal rule and is defined as the area below the line formed by joining total symptom scores on each daily diary card from day 0 through day 7, 14, and 21, respectively. The AUCs will be rescaled by time in order to provide results on a symptom scale from 0 to 39. This will be done by dividing the AUC by 7, 14, or 21, respectively, corresponding to the number of trapezoids created from daily diary cards between day 0 and the last day considered in the calculation (i.e., day 7, day 14, and day 21).

Participants who die or are hospitalized in the time interval being considered (through day 7, day 14, or day 21, respectively) will be ranked as worse (i.e., worse severity) than those alive and never hospitalized in worsening rank order. Programmatically, participants who die in the time interval will be assigned an AUC (severity score) of 42 (worst rank) regardless of when the death occurred in the interval, participants who are alive but remain hospitalized at last day of the interval will be assigned an AUC (severity score) of 41 (second worst rank), and participants who are alive but are no longer hospitalized on the last day of the interval will be assigned an AUC (severity score) of 40 (the third worst rank).

Participants who have incomplete diary cards for reasons other than hospitalization or death will be addressed in the same manner as the analyses of COVID-19 severity through day 28, outlined in the above section of the SAP.

Oxygen Saturation

The primary analysis will be repeated without adjustment for baseline oxygen saturation level. In addition, the absolute difference in proportion of participants with oxygen saturation ≥ 96% will be calculated at each measurement time, with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

For analyses based on interim data (e.g. DSMB reviews), the proportion of participants with oxygen saturation ≥ 96% will also be compared using log-binomial models fit separately at each time point. If at a given time point there are zero events in either arm, a p-value from Fisher's exact test will be provided instead. If there are zero events in both arms, then this will be stated and no formal statistical inferential analyses will be undertaken.

Subgroup Analyses

Duration of Clinical Symptoms

In phase III, to evaluate the effect of the investigational agent on symptom duration in specific populations (address secondary objective 8), secondary outcome 4 will be assessed among different subgroups. These will also be conducted for the supportive outcome of time to two consecutive days of resolution of all symptoms to "absent". Descriptive analyses for the following subgroups will be considered. A separate analysis plan for multivariate/personalized-medicine type analyses across subgroups will be developed at a later time.

Pre-specified subgroups of interest include:

1) Sex (Male sex at birth, female sex at birth)
2) Race (white, non-white)
3) Ethnicity (Hispanic, non-Hispanic)
4) Age Group (<60, ≥60)
5) Calendar days from first symptom associated with COVID-19 to start of investigational agent/comparator intervention Stratification (≤ 5 days, > 5 days)
6) Country (U.S., non-U.S.)
7) SARS-CoV-2 Variant (if available: categories will be determined based on prevalence of variants identified in testing)

5.2.2 Secondary Virology

The schedule of evaluations in protocol version 7.0 indicates that only NP swabs will collected in both phase II and phase III, and therefore only analyses of SARS-CoV-2 RNA from NP swabs are outlined below. Some agents may have completed enrollment in phase II prior to implementing protocol version 7.0, and therefore may have additional specimens collected for SARS-CoV-2 RNA testing. If analyses of these additional specimens are pursued, then the approach will be as defined in the relevant previous version of the SAP.

Analysis Population

The analyses of the virology objectives will include all randomized participants who started an investigational agent or the concurrent comparator intervention, according to a modified intent-to-treat (mITT) approach (Treated Population). Participants who have protocol violations will be included in the analysis but the protocol violations will be documented and described.

Note: Participants who are randomized but do not start investigational agent or comparator intervention are, per protocol, not to be followed and will be replaced.

Analysis Methods

Quantification (< LLoQ versus ≥ LLoQ) of SARS-CoV-2 RNA at day 3 (this is a secondary outcome for the active-controlled phase 3 only)

Descriptive statistics (number and percentage) will be used to describe the proportion of participants with SARS-CoV-2 RNA < LLoQ from staff-collected NP swabs at entry and day 3.

The proportion of participants with SARS-CoV-2 RNA < LLoQ day 3 will be compared between randomized arms using log-binominal regression for repeated binary measurements with log-link. The model will include a main effect for treatment and will adjust for baseline (day 0) log-10 transformed SARS-CoV-2 RNA level. The estimated adjusted relative risk of having RNA < LLoQ (and associated 95% CI and two sided p-value) will be obtained by taking the exponential of the treatment parameter estimate (and associated 95% CI). In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead.

In this analysis, baseline SARS-CoV-2 RNA values will be imputed if the level is < LLoQ as outlined in section 4.1. It is not expected that a high proportion of baseline results will be < LLoQ; however, in the event that there is a non-negligible proportion of baseline results < LLoQ (defined as 10% or more of baseline results < LLoQ), an additional variable will be added to the model that will indicate whether the baseline result was above or below the LLoQ (included programmatically as “0” if above LLoQ, and “1” if below LLoQ). Missing data are assumed to be missing completely at random (MCAR) and will be ignored in these analyses; however, sensitivity analyses will address possible informative missingness (see below).

Level (Quantitative) of SARS-CoV-2 RNA

Descriptive statistics will be used to describe the levels of SARS-CoV-2 RNA at each scheduled measurement time for staff-collected NP swabs.

Non-parametric Wilcoxon rank-sum tests with a 5% type I error rate will compare SARS-CoV-2 RNA level (continuous) between randomized arms, separately at each post-entry study day; results below the limit of detection will be imputed as the lowest rank and values above the limit of detection but below the LLoQ will be imputed as the second lowest rank.

Missing data in analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

AUC of SARS-CoV-2 RNA

In phase II only, levels of log-10 transformed SARS-CoV-2 RNA, measured from NP swabs will be analyzed using participant-specific AUCs. In this analysis, the AUC is defined as the area below the line formed by joining measured values at each successive measurement time and

above the lower limit of quantification of the assay, calculated using trapezoidal rule. Programmatically, the trapezoidal rule will be applied to the following values: max[0, $\log_{10}$(RNA)-$\log_{10}$(LLoQ)], obtained at the scheduled measurement times between and including day 0 and day 14.

Missing values with preceding and succeeding values will be ignored, which is equivalent to linearly interpolating the RNA levels from preceding and succeeding values. Missing values with no succeeding values will be imputed using linear imputation assuming that the RNA level at day 14 equals the LLoQ (as it is anticipated that nearly everyone will clear virus over 14 days). If the day 0 result is missing then the participant will be excluded from analysis. The participant-specific AUCs will be compared between randomized arms using a two-sided Wilcoxon test with 5% type I error rate.

Missing data in the AUC analysis of continuous SARS-CoV-2 RNA levels are assumed to be missing completely at random (MCAR) and will be ignored in analysis.

Sensitivity Analyses

The following sensitivity analyses are included to evaluate impact of different assumptions on the inference of the virology outcomes.

All Virology Outcomes

1) Repeat primary analysis, but restrict analysis population to exclude those with SARS-CoV-2 RNA < LLoQ at Day 0. This model will adjust for baseline log-10 transformed SARS-CoV-2 RNA level.

Dichotomous Virology Outcomes

1) Repeat primary analysis, but impute missing data in the following manner (ignores missingness due to hospitalization and death):
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used
2) Repeat primary analysis, but impute missing data in the following manner (special considerations for missingness due to hospitalization and death):
    - For missingness due hospitalization or death, participants with missing SARS-CoV-2 results will have their values imputed as ≥ LLoQ.
    - For non-monotonic missingness, participants with missing SARS-CoV-2 results will have their values imputed as < LLoQ if the preceding and succeeding results are < LLoQ, otherwise the results will be imputed as ≥ LLoQ.
    - For monotonic missingness, inverse probability weighted GEE will be used.

Supportive Analysis

The dichotomous virology analysis will be repeated without adjustment for baseline (Day 0) SARS-CoV-2 RNA level. In addition, the absolute difference in proportion of participants with RNA < LLoQ will be calculated at each measurement time; with associated 95% confidence intervals (calculated using the normal approximation to the binomial distribution).

5.3 Exploratory Analyses

5.3.1 New SARS-CoV-2 among Household Contacts

The analysis of household contacts will be restricted to the subset of randomized participants in the Treated Population who reported that they share indoor living space or housekeeping space with someone.

New SARS-CoV-2 positivity among household contacts through day 28 will be analyzed in the following manner. The proportion of participants with a household contact that tests positive for SARS-CoV-2 after the participant initiates study investigational agent or concurrent comparator intervention through day 28, will be estimated and compared between randomized arms using log-binomial regression, with log link, in order to obtain a risk ratio estimate; the model will include a main effect for randomized arm. In the event the log-binomial regression model fails to converge, a Poisson regression model with robust variance and log-link will be used instead. Missing data will be considered missing completely at random in analysis.

The same analysis approach will be used to compare the proportion of participants with a household contact that tests positive for SARS-CoV-2 or has COVID-19 symptoms after the participant initiates study investigational agent or concurrent comparator intervention through day 28.

Analysis of new SARS-CoV-2 positivity, and new SARS-CoV-2 positivity or COVID-19 symptoms, among household contacts through week 24 will be analyzed as in the same way as above for these outcomes through day 28.

5.3.2 Hospitalization Course

Analyses of clinical outcomes among those hospitalized will include all randomized participants who started an investigational agent or the concurrent comparator intervention who were also hospitalized. The analyses will be limited to descriptive summaries by randomized arm, as these analyses are restricted to participants who were hospitalized and so are not randomized comparisons.

Duration of hospitalization and duration of ICU admission will be summarized with continuous descriptive statistics. Duration of hospitalization/ICU through day 28 will be calculated as the difference between the date of discharge and the date of admission; the duration will be truncated at Day 28, if the participant is still hospitalized at Day 28. If data on discharge dates occurring after Day 28 are complete at the time of analysis of the Day 28 data, an additional descriptive analysis of durations for hospitalizations starting on or before Day 28 will be undertaken. The proportion of participants with ICU admission, among those hospitalized, will be summarized with

frequencies and percentages. The worst clinical status (ordinal outcome) will be summarized with frequencies and percentages. Descriptive summaries of use of remdesivir and dexamethasone, and other approved medications for treatment of COVID-19 used during hospitalization will also be included.

This analysis will be done through day 28 and separately through week 72.

5.3.3 Resistance Mutations

Analyses addressing the emergence of new resistance mutations will be outlined for each investigational agent in agent-specific SAP appendices based on information about resistance available at the time of completion of sequencing.

5.4 Interim Analysis Considerations

Interim analyses of the placebo-controlled superiority phase III evaluation of an agent was finished at the time of finalization of SAP version 7.0. The following from protocol version 7.0 describes the interim analysis considerations for the active-controlled non-inferiority phase III evaluation of an agent.

The two-sided 95% confidence interval mentioned above [see section 2.5.3 of the SAP] will be adjusted for the multiple interim analyses to preserve the confidence interval coverage to at least 95% (this is also referred to as using “repeated” confidence intervals).

The standard Lan and DeMets approach will be used to achieve this, incorporating an O’Brien and Fleming spending function. For simplicity, the information scale for the spending function will be determined as the proportion of the planned enrollment randomized to the investigational agent being evaluated at the time of the interim analysis. As an example, if in practice, the analyses were after exactly 25%, 50%, 75% and 100% of the planned enrollment, then the nominal confidence intervals used to assess efficacy would have coverage 99.9985% at the first analysis, 99.70% at the second analysis, 98.17% at the third analysis and 95.60% at the fourth analysis (these were obtained from PASS software). However, as the O’Brien and Fleming spending function is very conservative at early interim analyses, making stopping very difficult, for the assessment of inferiority of an investigational agent compared to the active comparator agent, an asymmetric approach will be used to reduce the level of evidence required for early stopping in the event that an investigational agent appears inferior to the active comparator agent. Specifically, if a nominal confidence interval with coverage of greater 99.9% at an early interim analysis is suggested by use of the O’Brien and Fleming spending function, then a nominal confidence interval with coverage of 99.9% will be used instead for assessing inferiority of the investigational agent.

The DSMB will also monitor the proportion hospitalized/dead in the active comparator arm as this key parameter, coupled with the non-inferiority margin, underpins the study design. The study is designed assuming that the underlying true proportion of participants on the active comparator agent is 2.3%. This is the proportion (32/1392) observed for high risk participants in the Regeneron COV-2067 trial for the agent (pooling across doses studied in that trial; FDA communication to DAIDS/NIAID). A 95% confidence interval for this proportion is (1.5%, 3.1%).

An assessment of non-inferiority in this study would be more difficult if the proportion of participants on the active comparator agent in this study is somewhat different from that in the Regeneron COV-2067 (e.g., somewhat outside of the range suggested by the confidence interval).

For example, this might arise if variants of SARS-CoV-2 are present in the study population which the active comparator agent is less effective against. Such an issue would undermine the use of a 3% non-inferiority margin in this study. It may however be addressed by focusing the non-inferiority assessment on the subpopulation in this study without such variants (assuming these have been identified), or in establishing superiority of the investigational agent in the overall study population. This may require a larger sample size to maintain power.

Another potential reason for a somewhat different proportion hospitalized/dead on the active comparator agent in this study versus that in the Regeneron COV-2067 study is that this study is likely to enroll in a number of different countries, whereas the Regeneron COV-2067 enrolled primarily in the United States. Aside from possible differences in circulating variants among countries, differences among countries in clinical practice and/or in the availability of hospital care might lead to differences in hospitalization/death rates. The DSMB will monitor descriptive results by country and provide guidance about countries with notably low or high rates of hospitalization/death.

6 Appendix 1: Algorithm for Handling Missing Symptom Evaluations for the Primary Phase II Symptom Outcome Measure.

The following algorithmic approach will be used to handle hospitalizations and deaths, as well as missing data, in constructing the TTE symptom-based outcome measure. The steps of the algorithmic approach will be undertaken in the following order:

a. If a participant has none of the targeted symptoms evaluated at any time during follow-up (including if due to the diary never being returned):

i. If the participant died on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved prior to death but will be retained in the risk set through to 28 days (programmatically, this is achieved by considering the participant censored after 27 days). [The underlying premise is that (if evaluations had been available) the participant had targeted symptoms that did not improve/ resolve through to death. Retention in the risk set through to 27 days provides for appropriate estimation of the cumulative proportion of the Treated Population who had a good outcome, i.e. symptoms improved/resolved for two consecutive days].

ii. If the participant was hospitalized on or before study day 28, then the participant will be assumed not to have had symptoms improved/resolved through to the day of hospital discharge and their follow-up will be censored at the day before hospital discharge (or at day 27 if earlier). [The underlying premise is that (if evaluations had been available) the participant had symptoms that did not improve/resolve through to admission to hospital and during hospitalization. Censoring at the day before hospital discharge assumes that the participant's subsequent unobserved symptom course would have been the same as other participants who were still at risk on the study day that discharge occurred].

iii. If the participant was not known to have died or been hospitalized, then their follow-up will be censored at day 0. [Censoring at day 0 assumes that their subsequent unobserved symptom course would have been the same as other participants in the Treated Population].

b. If a participant has one or more (but not all) targeted symptoms with no evaluations for all days from day 0 through day 28:

The TTE outcome measure for this participant will be evaluated based on the remaining targeted symptoms with missing data handled for those targeted symptoms as described below in subsection c. [In essence, this is assuming that if the participant had evaluated the unscored symptoms that they would have shown improvement/resolution for two consecutive days as the same time, or earlier, as the symptoms that they did score. With this assumption, using the available symptom data is considered preferable to alternative strategies of censoring their TTE at day 0 or assuming that the unscored symptoms never improved/resolved throughout follow-up with censoring at day 27].

c. If participant has an evaluation on day 0 and/or on days between day 1 and day 28 during follow-up on all targeted symptoms (or, per section b above, on a subset of targeted symptoms):

For each symptom having an evaluation on at least one day between day 0 and day 28 inclusive, programmatically values will be imputed for unobserved evaluations after death, for days in hospital, and for missing values as follows:

i. For days after death (and the day of death if no diary was completed that day), set all symptoms to "severe". This means that each symptom is never considered improved/ resolved unless this was achieved prior to death. For participants who did not achieve the event prior to death, the effect of this is to retain them in the risk set from death through to 28 days without meeting the symptom improvement/resolution criteria providing for appropriate estimation of the cumulative proportion of the Treated Population who had symptoms sufficiently improved/resolved throughout follow-up time.

ii. For days hospitalized (including day of admission if no diary was completed that day, and including the day of discharge if no diary was completed that day), set all symptoms to "severe" irrespective of whether or not the diary was completed. This means that each symptom is not considered improved/ resolved while a participant was hospitalized, but note that a participant could still have achieved the symptom outcome criteria prior to hospitalization.

iii. Impute a missing score for a symptom on day 0 as "mild". If also missing on day 1 or for a sequence of consecutive days from day 1 but with at least one score during follow-up, impute the missing values on day 1 through to the first available score as "mild". This means that the TTE criteria cannot be met during follow-up while a participant has a sequence of one or more missing values starting on day 0. The choice of imputing a missing value as "mild" on day 0 means that that symptom has to resolve to "absent" during follow-up before the TTE criteria can be met.

iv. For intermittent missingness during follow-up after day 0, impute a missing score for a symptom as the worst of (a) the last available value (actually provided by the participant or imputed due to hospitalization) before the missing value, and (b) the first available value (actually provided by the participant or imputed due to hospitalization) after the missing value, irrespective of the length of the sequence of missing values for the symptom. This gives potentially longer times until symptom improvement/resolution (compared with what might have occurred if the evaluations were available) if either of the preceding and succeeding values do not meet the criteria for improvement/ resolution, but potentially shorter times if both the preceding and succeeding values meet the criteria.

v. For monotonic missingness through to day 28 (i.e. a sequence of missing values during follow-up through to and including day 28 due to loss to follow-up, participant choice not to fully complete their diary, or an early day 28 clinic visit at which the diary is returned), censor the follow-up for this specific symptom at the last day that the relevant criterion for symptom improvement could have been met (this would be the day before the last diary entry for a given symptom, the day before the day of discharge, or the day before the day of withdrawal from the study during

hospitalization). This assumes that the censoring is non-informative about when the criterion would have been met if diaries had been fully completed.

The TTE outcome is then calculated as the first of two successive days meeting the symptom improvement/resolution criteria using the combined observed and imputed data for all symptoms with one or more evaluations observed during follow-up between day 0 and day 28, inclusive. In the event that the censoring due to monotonic missingness differs among targeted symptoms (e.g. because the participant stops completing the diary for one symptom earlier than for other symptoms), then the TTE outcome will be calculated using the available observed and imputed data, and censoring of the TTE outcome will be at the time of censoring of the symptom with the longest time to censoring.

7 Appendix 2: Statistical Considerations for BRII-198 + BRII-196

NOTE: Enrollment to BRII-198+BRII-196 started under protocol version 2 and continued through to protocol version 6.0. There were changes to the phase II primary virology and symptom outcomes measures in protocol version 3 from protocol version 2. No analyses comparing BRII-198+BRII-196 to placebo for the protocol version 2 phase II outcomes had been undertaken when protocol version 3 was implemented, and this SAP documents the intent that the phase II primary virology and symptom outcome measures in protocol version 3 (and continued in subsequent protocol versions) are primary using data from participants enrolled under all protocol versions. All participants enrolled to evaluate BRII-198+BRII-196 were randomized to active agent or placebo and so placebo is mentioned as the comparator intervention throughout this appendix.

7.1 Randomization Details

Phase III is stratified by time from symptom onset (≤ 5 days versus > 5 days).

7.2 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

2) Phase III only: New Grade 3 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 3 or higher AE is defined as: Grade 3 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 or 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

7.3 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BRII-196+BRII-198 or a placebo arm for an agent with follow up through to at least week 72).

8 Appendix 3: Statistical Considerations for AZD7442 IV

NOTE: AZD7442 IV is only being evaluated in this study in phase II with a placebo control.

8.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

8.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IV or a placebo arm for an agent with follow up through to at least week 72).

9 Appendix 4: Statistical Considerations for AZD7742 IM

NOTE: AZD7442 IM is only being evaluated in this study in phase II with a placebo control.

9.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment

9.2 Analysis Approaches

The secondary safety outcome measure specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for AZD7442 IM or a placebo arm for an agent with follow up through to at least week 72).

10 Appendix 5: Statistical Considerations for SNG001

10.1 Objectives

10.1.1 Secondary Objectives

1) Phase II: To evaluate SNG001 adherence compared to placebo for SNG001 over the 14-day treatment period.

2) Phase III: To determine whether SNG001 reduces hospitalization or death through study day 28 among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

3) Phase III: To determine whether SNG001 reduces duration of targeted COVID-19-associated symptoms through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

4) Phase III: To determine whether SNG001 reduces COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

5) Phase III: To determine whether SNG001 reduces progression of one or more COVID-19-associated symptoms to a worse status than recorded in the study diary at entry, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

6) Phase III: To determine whether SNG001 increases proportion of individuals with pulse oximetry measurement of ≥ 96% through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

7) Phase III: To determine whether SNG001 reduces the time to sustained symptom resolution through study day 28, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

8) Phase III: To determine whether SNG001 prevents the composite endpoint of either hospitalization due to any cause or death due to any cause through study week 72, among individuals in the subgroup who report moderate or severe shortness of breath or

difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

9) Phase III: To determine whether SNG001 prevents the composite endpoint of hospitalization or death through stay day 28, excluding hospitalizations that are determined to be unrelated to COVID-19, among individuals in the subgroup who report moderate or severe shortness of breath or difficulty breathing at day 0, and in the subgroup who report severe shortness of breath or difficulty breathing at day 0.

10.1.2 Exploratory Objectives

1) Phase II and III: To determine whether SNG001 reduces severity of shortness of breath or difficulty breathing through study day 28.

2) Phase II: To determine whether SNG001 reduces a COVID-19 Severity Ranking scale based on COVID-19-associated symptom burden (severity and duration), hospitalization, and death, through study day 28 among individuals who report moderate to severe shortness of breath or difficulty breathing at day 0.

10.2 Outcome Measures

10.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of SNG001 or placebo for SNG001.
2) Phase II only: Percentage of the 14 doses of SNG001 or placebo for SNG001 that are missed, defined as the number of missed doses divided by 14.

10.2.2 Other Outcome Measures

1) Phase II: Area under the curve of *shortness of breath or difficulty breathing* symptom severity over time from the participant's diary from day 0 to day 28.

   For participants who are alive at 28 days and not previously hospitalized, symptom severity on a given day is defined as the sum of scores for the *shortness of breath or difficulty breathing* symptom in the participant's study diary (each individual symptom is scored from 0 to 3). Participants who are hospitalized or who die during follow-up through 28 days will be ranked as worse than those alive and never hospitalized as follows (in worsening rank order): alive and not hospitalized at 28 days; hospitalized but alive at 28 days; and died at or before 28 days.

10.3 Analysis Approaches

10.3.1 Secondary Analyses

Secondary analyses of adherence will be restricted to those who received at least one dose of SNG001 or placebo for SNG001, and will not include others in the pooled placebo group as adherence is only assessed in those who took SNG001 or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least one dose of SNG001 or placebo for SNG001, and will be compared between arms using binary regression, in

a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

The secondary objectives addressing analyses among people who reported moderate or severe shortness of breath or difficulty breathing at day 0, and among people who reported severe shortness of breath or difficulty breathing at day 0, will be undertaken in the same manner as the analyses of this outcomes among the overall study population.

10.3.2 Exploratory Analyses

Exploratory analyses will compare the AUC for shortness of breath or difficulty breathing between arms; this analysis will include all participants in the Treated Population (i.e. will include the full pooled placebo group). The AUC will be calculated using the same methods as the overall COVID-19 symptom severity ranking (secondary outcome) and will be compared between arms using a two-sided Wilcoxon test with a 5% type I error rate.

To address SNG001-specific exploratory objective 2, the COVID-19 severity ranking outcome will compared between arms using the same methods outlined for the secondary analysis of this outcome measure, but restricted to be among those with moderate to severe shortness of breath or difficult breathing at day 0,

11 Appendix 6: Statistical Considerations for Camostat

Note: camostat is only being evaluated in this study in phase II with a placebo control.

11.1 Objectives

11.1.1 Secondary Objectives

1) Phase II: To evaluate camostat adherence compared to placebo for camostat over the 7-day treatment period.

11.1.2 Exploratory Objectives

1) Phase II: To explore the relationship between camostat adherence and study outcomes.

11.2 Outcome Measures

11.2.1 Secondary Outcome Measures

1) Phase II only: Number of missed doses of camostat or placebo for camostat.
2) Phase II only: Percentage of the 28 doses of camostat or placebo for camostat that are missed, defined as the number of missed doses divided by 28.

11.3 Analysis Approaches

Secondary analyses of adherence will be restricted to those who received at least one dose of camostat or placebo for camostat, and will not include others in the pooled placebo group as adherence is only assessed in those who took camostat or the matching placebo. Adherence will be evaluated by estimating the proportion of participants who missed at least four doses of camostat or placebo for camostat, and will be compared between arms using binary regression, in a similar manner as the primary safety outcomes. The percentage of missed doses will be compared between arms using a two-sided Wilcoxon test with 5% type-I error rate.

Analyses to address exploratory objective 1 will be developed in future analysis plans, depending on the results of analyses addressing the primary and secondary objectives.

12 Appendix 7: Statistical Considerations for SAB-185

There are two doses of SAB-185 under consideration in phase II (3,840 Units/kg dose group or the 10,240 Units/kg dose group), each of which will be considered a separate agent group in analysis.

There are no agent-specific objectives or outcomes measures.

However, there are key special study design and analysis considerations for the phase III evaluation of SAB-185 (3,840 Units/kg). These are described in this appendix based on material presented in sections 3.4 and 10 of protocol version 8.0.

12.1 Study Design and Analysis Population Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185

Phase III evaluation of SAB-185 was initiated under protocol version 7.0 in a non-inferiority comparison of SAB-185 to an active control of casirivimab plus imdevimab. While enrollment was ongoing, the Omicron variant of SAR-CoV-2 became highly prevalent. In vitro data suggested that casirivimab plus imdevimab would be ineffective against this variant, and FDA authorization for emergency use of this regimen for treatment of COVID-19 was withdrawn due to non-susceptible SARS-CoV-2 variants (such as Omicron). Because of this, enrollment into the study was paused pending development of protocol version 8.0, which replaces the non-inferiority evaluation of investigational agents compared to casirivimab plus imdevimab with a placebo-controlled superiority design allowing for the additional use of standard of care (SOC) treatments, if available, in both arms.

Over 700 participants were enrolled under protocol version 7.0 and randomized to SAB-185 or casirivimab plus imdevimab. These participants can be divided into two subpopulations: (1) participants who were definitely or very likely infected with the Omicron SARS-CoV-2 variant (“Omicron Subpopulation”); and (2) participants who were definitely not, or likely not infected with the Omicron variant (“Non-Omicron Subpopulation”). Based on the details in section 10.1.1 of protocol version 8.0, these two subpopulations were defined in more detail in version 9.0 of this SAP as follows:

- The “Omicron Subpopulation” enrolled under protocol version 7.0 is defined as (1) all participants randomized under protocol version 7.0 who were infected with the Omicron variant as identified on sequencing of an NP sample obtained on day 0, plus (2) all participants randomized under protocol version 7.0 on or after December 26, 2021, who do not have variant information available from a sample obtained on day 0. The second of these two groups of participants are assumed very likely to be infected with the Omicron variant on the basis that prevalence of the Omicron variant in the U.S. was estimated by the CDC to be 89.2% for the week ending January 1, 2022 (and starting December 26, 2021), 95.2% for the week ending January 8, 2022, 97.9% for the week ending January 15, 2022, and 98.9% for the week ending January 22, 2022, during which enrollment under protocol version 7.0 was stopped [COVID Data Tracker. United States Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#variant-proportions. Accessed February 11, 2022].

- The “Non-Omicron Subpopulation” enrolled under protocol version 7.0 is defined as all participants enrolled under protocol version 7.0 excluding those in the “Omicron Subpopulation.” It therefore includes all participants randomized on or before December 25, 2021, who did not have the Omicron variant identified in sequencing of an NP sample obtained on day 0 (including those with no sequence available from an NP sample obtained on day 0). It also includes all participants randomized on or after December 26, 2021 who had a non-Omicron variant identified in sequencing of an NP sample obtained on day 0.

These definitions are updated in version 10.0 of the SAP to be based on the timing of emergence of the Omicron variant within the study population. This is to reduce the risk of misclassification between the two subpopulations and hence improve the value of information from the study, particularly for secondary outcomes, despite its early termination for operational futility. The study team reviewed the available variant information for study participants and noted (a) that the completeness of variant information would be increased substantially by including variants determined in samples obtained during follow-up and not just those obtained on day 0 (and that variant determinations were consistent in all participants who had variant information from multiple samples); and (b) that the Omicron variant became dominant in the study population from December 15, 2021 onwards. The study team therefore decided that the following adjustments to the definitions above should made:

- Variant information from any sample (i.e., not just from samples obtained on day 0) could be used to assign a participant to the Omicron versus non-Omicron Subpopulations.

- For participants without variant information, those randomized under protocol version 7.0 on or after December 15, 2021 would be assigned to the Omicron Subpopulation, and those randomized on or before December 14, 2021 would be assigned to the non-Omicron Subpopulation.

Based on in vitro data, the combination of casirivimab plus imdevimab is thought to have no effect on hospitalization/death in the Omicron Subpopulation and so is considered functionally to be a placebo from an efficacy perspective. Therefore, for the purposes of evaluating the superiority of SAB-185 versus placebo under this version of the protocol, the Randomized Population will be comprised of the Omicron Subpopulation enrolled under protocol version 7.0 as well as the population enrolled under protocol version 8.0 that is randomized to SAB-185 or its appropriate placebo control group. The planned sample size for this Randomized Population combining participants in the Omicron Subpopulation enrolled under protocol version 7.0 and participants enrolled under version 8.0 of the protocol is 1200 participants. The Treated Population (modified intent-to treat [mITT] population) for evaluating SAB-185 is this Randomized Population after excluding any participants who did not receive study product (i.e. excluding those who did not start SAB-185 or casirivimab plus imdevimab if enrolled under protocol version 7.0, and excluding those who did not start SAB-185 or placebo if enrolled under protocol version 8.0). Note that enrollment to the phase III evaluation of SAB-185 was terminated due to operational futility prior to the opening of randomization to SAB-185 or placebo, and in the Omicron Subpopulation, the comparison in practice will be between participants randomized to SAB-185 versus casirivimab plus imdevimab.

The Non-Omicron Subpopulation enrolled under protocol version 7.0 is not part of the phase III placebo-controlled superiority evaluation of SAB-185. See section 12.4 of this appendix for analysis considerations for this population.

12.2 Additional Analysis Considerations Specific to the Phase III Placebo-Controlled Superiority Evaluation of SAB-185

Data from the Treated Population defined in section 12.1 of this appendix will be used to evaluate the efficacy and safety of SAB-185. As noted above, because the Phase III evaluation of SAB-185 was terminated early due to operational futility and prior to any participants being randomized to receive placebo, this evaluation now involves a comparison of participants in the Treated Population randomized to SAB-185 versus casirivimab plus imdevimab. These analyses will follow the analysis plans in this SAP for the phase III placebo-controlled superiority evaluation of an investigational agent.

Descriptive summaries of temporal patterns of variants will be included as part of the Phase III analyses of SAB-185.

12.3 Analysis Considerations for the Non-Omicron Subpopulation Enrolled under Protocol Version 7.0

Follow-up of the Non-Omicron Subpopulation enrolled under protocol version 7.0 will continue per protocol. In this subpopulation, the combination of casirivimab plus imdevimab is expected to be effective. Analysis of outcomes from the Non-Omicron Subpopulation will be undertaken separately from analyses involving the Omicron Subpopulation, following the plans laid out in protocol version 7.0 and described in this SAP for the active-controlled non-inferiority Phase 3 trial. It is recognized that there will be limited power to assess non-inferiority with respect to the hospitalization/death primary outcome measure.

12.4 Data and Safety Monitoring for the Evaluation of SAB-185 under Protocol Version 8.0

[At the time of preparing this version of the SAP, all participants randomized in the superiority phase III evaluation of SAB-185 have completed study treatment and the day 28 intensive follow-up. Text in this section therefore provides a record of the monitoring plan that was in place for this part of the study].

In addition to the details regarding data and safety monitoring laid out in the Master Protocol, the DSMB may consider results from the “Non-Omicron Subpopulation” enrolled under protocol version 7.0 to guide their recommendations, particularly regarding any safety issues or possible early termination of the placebo-controlled evaluation of SAB-185 based on lack of sufficient efficacy. For example, data suggesting that SAB-185 may be less effective than casirivimab plus imdevimab in the Non-Omicron Subpopulation might support a finding of lack of sufficient efficacy of SAB-185 versus placebo in the “Omicron Subpopulation”. Note, however, that a recommendation to terminate randomization in the phase III placebo-controlled superiority evaluation of SAB-185 (being conducted under protocol version 8.0) based on a finding of superiority of SAB-185 versus placebo should, in general, be based only on results from the Treated Population for this comparison defined in section 12.1 of this appendix.

13 Appendix 8: Statistical Considerations for BMS-986414 + BMS-986413

13.1 Secondary Outcome Measures

1) Phase II only: New Grade 2 or higher AE through week 72. [Supportive of Primary Objective 1]

   New Grade 2 or higher AE is defined as: Grade 2 or higher event that was new in onset or aggravated in severity or frequency from the baseline condition (i.e., Grade 1 at baseline escalates to Grade 2 or higher, or Grade 2 at baseline escalates to Grade 3 or higher, or Grade 3 at baseline escalates to Grade 4 or higher), following the start of study treatment.

13.2 Analysis Approaches

The secondary safety outcome measures specified in this appendix will be analyzed in the same manner as the primary and secondary safety outcomes defined in the main text of the SAP. The same analysis population will be considered, however, in phase II, the placebo control arm will be restricted to those who were randomized to a placebo that included follow-up through at least week 72 (i.e. will be restricted the those who received placebo for BMS-986414 + BMS-986413 or a placebo arm for an agent with follow up through to at least week 72).