STATISTICAL ANALYSIS PLAN FOR PART II

Protocol CM4620-204

A Randomized Double Blind, Placebo-Controlled Study of Auxora for the Treatment of Severe COVID-19 Pneumonia (CARDEA)

PREPARED BY:

NAME: **Jeffrey Zhang, PhD.**

POSITION: **Principle Statistician, Princeton**

VERSION: **Pharmatech Final 2.0; NCT04345614**

DATE: **17 August 2021**

APPROVAL FORM

Statistical Analysis Plan for Part II

Protocol No. CM4620-204

Protocol Title A Randomized Double Blind, Placebo-Controlled Study of Auxora for the Treatment of Severe COVID-19 Pneumonia (CARDEA)

Version Final 2.0

Date 17 August 2021

Author Jeffrey Zhang, PhD,
Principle Statistician, Princeton Pharmatech LLC

APPROVED BY:

Aug-17-2021

Jeffrey Zhang, PhD
Principle Statistician, Princeton Pharmatech LLC

Date

17-Aug-2021

Sudarshan Hebbar, MD
Chief Medical Officer, CalciMedica, Inc.

Date

DOCUMENT HISTORY - CHANGES COMPARED TO PREVIOUS FINAL VERSION OF SAP

Date	Outcome to update	Reason for update	Section and title impacted (Current)
Mar-31-2021	First version	Creation of final version	NA
Aug-15-2021	Amendment 1 (Second Version)	Clarification of the data collection period for the time to recovery per FDA comment	5.1 Efficacy endpoint
		Remove Completer Analysis Set. The completer analysis set analysis is changed to as treated analysis using safety analysis set	6 Analysis Populations
		Add additional analysis for Concomitant Medications of Vasopressors, Corticosteroids, Antivirals And Anticoagulants	8.3.3 Concomitant Medications
		Add additional sensitivity analysis for the prohibited immunosuppressive medications or immunotherapies and convalescent plasma treatment based on hypothesis censoring rule per FDA comment	8.4.3 Sensitivity Analyses
		Add additional supplementary analysis for including the baseline imputed PaO_2/FiO_2 of ≥ 201 stratification group in the analysis set per FDA comment.	8.4.4 and 8.5.4 Supplementary Analyses
		Clarification of the censoring rule details	8.5 Secondary Endpoint Analysis

Table of Contents

Table of Tables

List of Terms and Abbreviations

Abbreviation	Definition
AE	Adverse event
BLQ	Below the limit of quantification
BMI	Body mass index
CI	Confidence interval
COVID-19	Disease from infection with coronavirus 2019 or SARS-CoV-2
CM	Concomitant medication
CS	Clinically significant
DP	Decimal places
eCRF	Electronic data collection form
ECMO	Extracorporeal membrane oxygenation
IDMC	Independent data monitoring committee
LOCF	Last observation carried forward
MedDRA	Independent safety review committee
MMRM	Mixed-effect model repeated measure
PCS	Potentially clinically significant
PK	Pharmacokinetic(s)
PT	Preferred term
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SFISD	Start of first infusion of study drug
SOC	System organ class
TEAE	Treatment-emergent adverse event
TFLs	Tables, figures and listings
ULN	Upper limit of normal
WHO	World health organization

1 INTRODUCTION

This statistical analysis plan (SAP) is based on the CM4620-204 Protocol version 8.1 dated 31 March 2021. The plan covers statistical analysis of efficacy and safety data to assess the efficacy and safety of Auxora compared to Placebo in patients with severe COVID-19 Pneumonia. The SAP is for the analysis of Part 2 of the CM4620-204 study.

2 STUDY OBJECTIVES

2.1 Primary Objective

To assess the clinical efficacy of Auxora in patients with severe COVID-19 pneumonia

2.2 Secondary Objectives

- To assess the safety and tolerability of Auxora in patients with severe COVID-19 pneumonia
- To assess the pharmacokinetic profile of Auxora in patients with severe COVID-19 pneumonia

3 STUDY DESIGN

3.1 Overall Study Design

Part 2 is a randomized, double blind, placebo-controlled study in which initially up to 400 patients receiving supplemental oxygen at Screening, and who meet all of the inclusion criteria and none of the exclusion criteria, were to be randomized 1:1 to receive Auxora plus standard of care or Placebo plus standard of care. After a blinded analysis of the mechanical ventilation and death rate in the subgroup of patients with a baseline imputed PaO_2/FiO_2 >200 who were randomized into the study, the number of patients in this subgroup were capped at 26. The study sample size remained 320 for patients with a baseline imputed PaO_2/FiO_2 ≤200. When enrolling both subgroups, patients were stratified to ensure balanced randomization between the Auxora and Placebo arms.

The dose of Auxora will be 2.0 mg/kg (1.25 mL/kg) administered at 0 hour, and then 1.6 mg/kg (1 mL/kg) at 24 hours and 1.6 mg/kg (1 mL/kg) at 48 hours from the SFISD. The dose of Placebo will be 1.25 mL/kg administered at 0 hour and then 1 mL/kg at 24 hours and 1 mL/kg at 48 hours from the SFISD. The SFISD should occur within 12 hours of the patient or LAR providing informed consent. The dosing will be based on actual body weight obtained at the time of hospitalization or screening for the study. As described in the pharmacy manual, there will be an upper limit of the absolute dose (volume) of Auxora and volume of Placebo that will be administered for patients weighing more than 125kg.

A study physician or appropriately trained delegate will perform assessments at Screening, immediately prior to the SFISD, and immediately prior to each subsequent infusion. At 72 hours after the SFISD, the patient will be assessed every 24 hours (±4 hours) until 240 hours after the SFISD, then q48 hours until Day 30 after the SFISD, or until discharge if earlier. Patients who are discharged before Day 25 after the SFISD will be followed-up at Day 30 (±5 days) for a safety and mortality assessment. Patients who are discharged before Day 55 after the SFISD will be followed up at Day 60 (±5 days) for a safety and mortality assessment.

3.2 Schedule of Study Procedures

Table 1 Schedule of Assessments

	Screen	If Eligible, Baseline Assessment	R	0 hour	24 (±2) hours	48 (±2) hours	72 (±2) hours	96, 144, 192, 240 hours (±4 hours)	120, 168, 216 hours (±4 hours)	Days 12-28[g] (±4 hours)	Days 30, 60[e] (±5 days)
Informed Consent	X										
Daily Assessment - Ordinal Scale	X				X	X	X	X	X	X	
Weight and Height	X										
Vital Signs[a] including SpO_2 and FiO_2	X			X[f]	X	X	X	X	X	X	
Arterial Blood Gas[b]	X			X	X	X	X	X	X	X	
Pathogen Testing-COVID-19 diagnosis	X										
CXR or CT scan of the lungs	X										
Influenza A, B and Respiratory Panel[c]	X										
Daily Laboratory Monitoring[i]		X			X	X	X	X	X		
PK Analysis[h]					X	X	X				
Q72 hour Laboratory Monitoring[i]		X					X	X	X		
AE/SAE Assessment				X	X	X	X	X	X	X	X
Concomitant Medications				X	X	X	X	X	X	X	
Randomize patient			X								
Study Drug Administration				X	X	X					
IL-6 and IL-2R[d]		X						X[d]			

Hospitalized patients will complete all assessments. Patients will complete the Day 12 assessment if discharged home on oxygen prior today 12. All patients will complete the Day 30 assessment.

a. Vital Signs will include temperature, heart rate, respiratory rate, systolic blood pressure, diastolic blood pressure, and SpO_2 by pulse oximetry. For the SpO_2, both the values at the time of the assessment and the lowest noted from the time of the last assessment shall be recorded. The FiO_2 shall be recorded at the time of each SpO_2 recording.
b. Arterial blood gas results obtained as standard of care shall be recorded.
c. If performed as local standard of care.
d. Blood samples for IL-6 and IL-2R will be shipped to a reference lab and results may not be available to the clinical team while managing the patient. The IL-6 and IL-2R will be obtained prior to randomization and again at 96 hours. The IL-6 and IL-2R may be performed on blood drawn in the previous 12 hours.
e. Day 30 will be phone call assessment for patients who were discharged prior to Day 25. ; Day 60 will be a phone call assessment for patients discharged prior to Day 55
f. 0-hour vital signs required only if screening vital signs were measured >6 hours prior to SFISD.
g. Patients discharged prior to Day 12 with the need for supplemental oxygen will be contacted on Day 12 to determine if they are still receiving supplemental oxygen.
h. PK samples will be drawn at selected sites. Draw blood for PK analysis together with daily laboratory monitoring. The PK specimens may be obtained from blood drawn in the previous 12 hours. The time of the draw must be recorded.
i. Daily and Q72 hour laboratory monitoring may be performed on blood drawn in the previous 12 hours. Q 72 hours labs will

be performed at 72, 144 and 216 hours

4 SAMPLE SIZE

In the initial design of Part 2 of the study, 320 patients in the baseline imputed PaO_2/FiO_2 ≤200 subgroup and 80 patients in the baseline imputed PaO_2/FiO_2 >200 subgroup were to have been randomized. A constant recovery rate ratio of 1.43 was assumed for patients receiving Auxora compared to Placebo in both subgroups. In order to provide power of approximate 90% to detect a difference in the recovery rate ratio of approximately1.43 by a stratified 2-sided log-rank test at an overall 0.05 alpha level, given the 1:1 randomization, it was determined that the study would require 330 recovery events. With the study length of 30 days, the assessment of time to recovery required a sample size of 400 patients.

The enrollment of patients in the subgroup with a baseline imputed PaO_2/ FiO_2 >200 was stopped early because a blinded analysis of the data from the first 26 patients in the subgroup randomized into the study showed a low rate of mechanical ventilation or death. This finding was consistent with Part 1 of the study where no patients with a baseline imputed PaO_2/FiO_2 >200 required mechanical ventilation or died.

The Part 2 study sample size was then reevaluated given that the primary endpoint would focus on the subgroup of patients with a baseline imputed PaO_2/FiO_2 ≤200. A two group log-rank test with a 0.05 two-sided significance level would have 90% power to detect a difference in the recovery rate ratio of approximately 1.49 in the 320 patients with a baseline imputed PaO_2/FiO_2 ≤200 who were randomized 1:1 to Auxora or Placebo.

To further ensure that the sample size of 320 patients in the baseline imputed PaO_2/FiO_2 ≤200 subgroup is an adequate sample size, a sample size re-estimation procedure will be applied when the first 70 patients in the baseline imputed PaO_2/FiO_2 ≤200 subgroup have completed or discontinued from the study. The IDMC will perform the procedure and will make a nonbinding recommendation to CalciMedica to increase or not increase the study sample size. No ongoing study patients will be included in this sample size re-estimation procedure. The sample size will be re-estimated to provide a conditional power of 90%, based on the evaluation of the treatment efficacy. The sample size re-estimation procedure will not allow for a reduction in the planned sample size of 320 for the baseline imputed PaO_2/FiO_2 ≤200 subgroup. If the study randomizes 600 patients in the baseline imputed PaO_2/FiO_2 ≤200 subgroup it will provide 90% power if the recovery rate ratio is 1.34.

5 STUDY ENDPOINTS

5.1 Efficacy Endpoints

5.1.1 Primary Endpoint

Time to recovery through Day 60 visit

5.1.2 Key Secondary Endpoint

All-cause mortality at Day 60 visit

5.1.3 Supportive Secondary Endpoints

- All-cause mortality at Day 30 visit
- Proportion of patients requiring invasive mechanical ventilation or dying during the 60 days from the SFISD
- Proportion of patients requiring invasive mechanical ventilation during the 60 days from the SFISD
- Differences in outcomes as measured by an 8-point ordinal scale at Day 12
- Differences in outcomes as measured by an 8-point ordinal scale at Day 30
- Number of Days in the Hospital
- Number of Days in the ICU

The primary analyses of the efficacy endpoints will be based on the patients in the baseline imputed $PaO_2/FiO_2 \leq 200$ subgroup.

5.2 Safety Endpoints

- Vital signs measurements
- Laboratory measurements
- Concomitant medications
- Treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs)

5.3 Pharmacokinetic Endpoint

Not applicable for this SAP.

5.4 Pharmacodynamic Endpoint

Not applicable for this SAP.

5.5 Health-economics Endpoint

Not applicable for this SAP.

6 ANALYSIS POPULATIONS

6.1 All-Subjects Analysis Set

The All-Subjects Analysis Set will contain information from all screened subjects, including those who did not meet the study entry criteria or did not receive a study treatment.

6.2 Efficacy Analysis Set

Efficacy analyses will be based on the ITT principle. The Efficacy Analysis Set will include data from all subjects randomly assigned to study treatment. All data will be included and no subjects excluded because of protocol violations. For the analysis of efficacy data, subjects will be included in the treatment group according to their randomly assigned treatment.

Unless otherwise noted, all efficacy analyses will be performed using the Efficacy Analysis Set of the baseline imputed $PaO_2/FiO_2 \leq 200$ subgroup. All Efficacy Tests data will be listed.

6.3 Safety Analysis Set

The Safety Analysis Set will include data from all subjects randomly assigned to study treatment who receive any amount of Auxora or Placebo. All data will be included and no subjects excluded because of protocol violations. For safety data analysis, subjects will be included in the treatment group according to the treatment they actually receive.

7 TIMINGS OF ANALYSES

7.1 Safety Review

After the first 50 patients were randomized in Part 2, an IDMC evaluated safety data from the study. The IDMC will make a suggestion whether to continue the study without changes, or consider terminating the study due to a safety issue. The IDMC will again review the safety data when 70 patients with a baseline imputed $PaO_2/FiO_2 \leq 200$ have completed the study, and finally when 200 patients with a baseline imputed $PaO_2/FiO_2 \leq 200$ are randomized in Part 2 of the study. Adhoc safety reviews can be requested by the IDMC or sponsor during the study.

7.2 Interim Sample Size Re-estimation and Futility Analysis

One unblinded interim sample size re-estimation was planned and was conducted by the IDMC when 70 patients in the baseline imputed $PaO_2/FiO_2 \leq 200$ subgroup had completed the study. No ongoing study patients were included in this sample size re-estimation procedure. The sample size was re-estimated based on the interim estimation of the treatment recovery rate. The interim estimation of the recovery rate was also used in the non-binding futility analysis.

7.3 Final Analysis

The final analysis will be conducted when all randomized patients, based on the planned or the re-estimated sample size, complete the study.

8 STATISTICAL ANALYSES

8.1 General Considerations

The statistical evaluation will be performed using SAS® Version 9.4 or later.

In general, descriptive statistical methods will be used to summarize the data from this study. Unless stated otherwise, the term "descriptive statistics" refers to number of patients (n), mean, median, standard deviation (SD), minimum and maximum for continuous data, and frequencies and percentages for categorical data. Safety and efficacy summaries will be performed on the efficacy and safety analysis sets, respectively.

8.1.1 Number of Digits to Report

Table 2 Number of Decimal Places (DP)

Statistic	Specification	Apply to
Minimum, maximum	Same number of DPs as the data provided in the datasets	All original, i.e., non-derived, data provided in the datasets
Mean, median, confidence intervals	One more DP than the raw data	All
SD, SE	Two more DP than the raw data	All
Percentages	1 DP	All
p-values	3 DP	All
Odds Ratio	2 DP	All
Hazard Ratio	3 DP	All

8.1.2 Significance Level and Confidence Interval

The statistical tests will be performed as two-sided with significance level of 5%. The confidence intervals will be determined with a confidence level of 95%.

8.1.3 Descriptive Statistics Values to Calculate

Where appropriate, variables will be summarized descriptively (frequency and percent will be summarized for categorical variables; mean, SD, median, minimum, and maximum will be presented for continuous variables) by strata, study visit, and by treatment group.

The denominator for the percentages will be the total number of subjects in the treatment group and Analysis Set being presented, unless otherwise specified (e.g., on some occasions, percentages may be calculated based on the total number of subjects with available data at a particular visit and/or time point).

8.1.4 Derived Variables

8.1.4.1 Definition(s) of Baseline(s)

Unless otherwise specified, the baseline values are the last available assessment before or at the SFISD. If multiple measurements are recorded on the same day, the last measurement recorded prior to the dose of study medication will be used as the baseline value. If these multiple measurements occur at the same time or time is not available, then the average of these measurements (for continuous data) or the worst among these measurements (for categorical data) will be considered as the baseline value. Baseline imputed PaO_2/FiO_2 is the worst value in the 24 hours before screening.

8.1.4.2 Study Time

For patients enrolled in the study, the first dose of study medication will be administered on study hour 0. The study time in hours for an event is defined as event date and time – date and time of SFISD. The study day is defined as event date – date of SFISD + 1 day.

8.1.4.3 Definition(s) of (Percent) Change from Baseline(s)

For numerical variables, change from baseline will be calculated as the post-baseline value minus the baseline value. If percent change from baseline is required, then percent change from baseline will be calculated as the change from baseline divided by the baseline value, multiplied by 100. If baseline value cannot be determined for a particular variable, the change from baseline and percent change from baseline will not be calculated.

8.1.4.4 Adverse Events

Treatment-emergent AEs (TEAEs) are defined as AEs that started or worsened in severity on or after the SFISD.

All adverse events will be coded from the reported term using the Medical Dictionary for Regulatory Activities (MedDRA) version 23.0 or the later version.

8.1.4.5 Concomitant Medications

All concomitant medications will be coded using the WHO drug dictionary (B3 WHO Drug Dictionary Enhanced – Sep 2019 or the later version).

8.1.4.6 Lab Values Below/above the Limit of Quantitation Deriving

The lab values below the limit of quantification (BLQ) are collected in the form like "< 2" in the clinical data base. In this case the numeric value of the lab is a missing value. The cut-off value '2' will be used to impute the missing value for this case. Similarly, for the lab values above the limit of quantification, the cut-off value will be used for the numeric value of the lab. Local labs are used in this study and all lab parameters will be standardized to common SI units.

8.1.5 Missing Data

Missing data will not be imputed in the summary of safety endpoints at each time point. The missing data imputation methods for efficacy analyses are described in Section 8.

Imputation of missing or incomplete dates will be performed on AEs and concomitant medications (procedures) conservatively for determining the timing relative to the dose of study product unless otherwise specified. Partial or missing dates will be listed as recorded in the electronic case report form (eCRF).

8.1.6 Pooling of Sites

Data collected from all sites will be pooled together for the analysis.

8.2 Subject Disposition

The number of patients who enter screening will be summarized, and the percentage of these patients who fail to meet entry criteria will be reported for total subjects. Screen failures will be summarized in total and by each reason for screen failure.

Subject disposition will be summarized for patients' completion status of study treatment, status of outcome/discharge, and corresponding discontinue reasons in tables for each treatment group and then total subjects. The data listing for subject discontinuation will be generated.

8.3 Demographic, Other Baseline Characteristics and Medication

8.3.1 Demographic and Baseline Characteristic

Patient demographic and baseline characteristics will be summarized by mean, SD, median, minimum, and maximum for continuous variables; and by counts and percentages for categorical variables. Summaries will be provided separately for each treatment group and then total subjects. The following subject demographic and baseline characteristic are summarized. The demographic variables consist of age, age category (18-39, 40-64, 65+), sex, ethnicity and race. Baseline characteristics include height, weight, body mass index (BMI), time from symptom onset to randomization in days, baseline scores on 8-point ordinal scale, baseline imputed PaO_2/FiO_2 and category ($\leq$200, $\geq$201), baseline d-dimer, baseline ferritin, and baseline CRP. Efficacy and Safety Analysis Sets will be used for the summaries.

Other baseline variables such as relevant medical history or other medically relevant criteria could also be included in the baseline tables.

The 8-point ordinal scale is defined as:

Scale	Description
1	Death
2	Hospitalized, requiring invasive mechanical ventilation or ECMO
3	Hospitalized, requiring noninvasive mechanical ventilation or high flow supplemental oxygen
4	Hospitalized, requiring low flow supplemental oxygen
5	Hospitalized, not requiring supplemental oxygen but requiring ongoing medical care
6	Hospitalized, not requiring supplemental oxygen or ongoing medical care
7	Discharged, requiring supplemental oxygen
8	Discharged, not requiring supplemental oxygen

8.3.2 Comorbidity

A summary table of the number and percentage of patients by the category of comorbidity will be produced for patients in the Efficacy Analysis Set. The summary will be provided by treatment groups and then by total treatment groups. Comorbidity will be sorted alphabetically by categories of comorbidity. The number and percentage of patients with none, one and two or more categories of comorbidities will also be summarized. Listings for comorbidity will also be provided.

8.3.3 Concomitant Medications

Concomitant medication is defined as any medication, other than study medication, which is taken on or after the start day of treatment. The concomitant medication data will be coded using World Health Organization Drug Dictionary (B3 WHO Drug DDE – Sep 2019 or the later version). The number and percentage of subjects taking concomitant medications will be summarized overall and by ATC level 2 terms, preferred drug names and treatment groups. A separate listing for concomitant medications will also be provided.

Concomitant Medications of Vasopressors, Corticosteroids, Antivirals And Anticoagulants for COVID-19 Newly Started on or After Day 1 of Study Treatment will be summarized by treatment group.

8.4 Primary Endpoint Analysis

8.4.1 Definition of Endpoint

Time to recovery through the Day 60 visit in the 8-point ordinal scale is the primary endpoint.

The primary estimand corresponding to the primary endpoint is defined as:

1. Treatment: 2.0 mg/kg (1.25 mL/kg) of Auxora administered at 0 hour and 1.6 mg/kg (1 mL/kg) administered at both 24 hours and 48 hours from the Start of the First Infusion of Auxora with local standard of care.
2. Population: Severe COVID-19 pneumonia patients with a baseline imputed PaO_2/FiO_2 ≤200.
3. Variable: Time to recovery defined as time from SFISD to the first time a patient achieves the scale of 6, 7 or 8 on the 8-point ordinal scale for at least 24 hours.
4. Population level summary: Median difference using Kaplan-Meier estimate

Intercurrent events under consideration: 1) prohibited immunosuppressive medications or immunotherapies. 2) convalescent plasma treatment, 3) treatment discontinuation and 4) death.

The null and alternative hypotheses are:

- Null hypothesis H0: Time to recovery curves are the same between the two treatment arms
- Alternative hypothesis H1: Time to recovery are different between the two treatment arms

8.4.2 Main Analytical Approach

Time to recovery through Day 60 visit will be displayed using a Kaplan-Meier estimate and will be compared between the 2 treatment groups using stratified log-rank test stratified by baseline imputed PaO_2/FiO_2 screening of ≤100 vs. >100. Prohibited immunosuppressive medications or immunotherapies, convalescent plasma treatment and treatment discontinuation will be ignored (treatment policy strategy). Death will be treated as not recovered and time to recovery will be censored at Day 65 (composite strategy). Patients, including withdrawal of consent patients, will be censored at the last ordinal scale assessment if no recovery event is observed during the study (administrative censoring). The trial level Type I error of 0.05 between the two tests of the primary endpoint analysis and the key secondary endpoint analysis is controlled using the Benjamini and Hochberg method.

8.4.3 Sensitivity Analyses

There are three planned sensitivity analyses that target the primary estimand. The first will be to relax the assumption that the baseline imputed PaO_2/FiO_2 is a confounding factor, so the analysis will be repeated but without stratification by the baseline imputed PaO_2/FiO_2 of ≤100 vs. >100. This analysis implies the baseline risk is the same across strata.

The second sensitivity analysis is based on a different administrative censoring rule and same analysis as the main analysis will be performed. Patients will be censored at the last ordinal scale assessment before or on the outcome visit if no recovery event is observed before or on the

outcome visit. This is to address the data collection limitation between the Day 30 outcome visit and Day 60 visit assessment related to the assessment of the primary endpoint.

The third sensitivity analysis is based on a different censoring rule for the prohibited immunosuppressive medications or immunotherapies and convalescent plasma treatment. In this sensitivity analysis, all data after the prohibited immunosuppressive medications or immunotherapies and convalescent plasma treatment will be excluded from the analysis. The time to recovery will be censored at the beginning of the prohibited immunosuppressive medications or immunotherapies or convalescent plasma treatment.

8.4.4 Supplementary Analyses

There are four planned supplementary analyses.

The first supplementary analysis changes the population level summary to the recovery rate ratio. The recovery rate ratio and the 95% CI will be estimated using a Cox proportional hazard model with treatment as the independent variable.

The second supplementary analysis changes the population level summary to the recovery rate ratio. The recovery rate ratio and the 95% CI will be estimated using a stratified Cox proportional hazard model with treatment as the independent variable and stratified by the baseline imputed PaO_2/FiO_2 of ≤100 vs. >100.

The third supplementary analysis changes the population to the safety analysis set and the same analysis as the main analysis will be performed except that treatment group will be based on actual treatment received.

The forth supplementary analysis changes the population to the efficacy analysis set including 26 subjects from baseline imputed PaO_2/FiO_2 of ≥ 201 stratification group and the same analysis as the main analysis will be performed.

8.5 Secondary Endpoint Analyses

8.5.1 Key Secondary Efficacy Analysis

All-cause mortality at Day 60 visit is the key secondary endpoint.

The primary estimand corresponding to the key secondary primary endpoint is defined as:

1. Treatment: 2.0 mg/kg (1.25 mL/kg) of Auxora administered at 0 hour and 1.6 mg/kg (1 mL/kg) administered at both 24 hours and 48 hours from the Start of the First Infusion of Auxora with local standard of care.
2. Population: Severe COVID-19 pneumonia patients with a baseline imputed PaO_2/FiO_2 ≤200.
3. Variable: Dying by the Day 60 study follow-up visit.

4. Population level summary: difference in proportion of dying

Intercurrent events under consideration: 1) prohibited immunosuppressive medications or immunotherapies. 2) convalescent plasma treatment, and 3) treatment discontinuation.

The null and alternative hypotheses are:

- Null hypothesis H0: Proportion of dying by Days 60 is the same between the two treatment arms
- Alternative hypothesis H1: Proportion of dying by Days 60 is different between the two treatment arms

8.5.2 Main Analytical Approach

Proportion and 95% CI All-cause mortality at Day 60 visit will be displayed using a Clopper-Pearson interval and will be compared between the 2 treatment groups using a Cochran-Mantel-Haenszel test stratified by the baseline imputed PaO_2/FiO_2 of ≤100 vs. >100. Prohibited immunosuppressive medications or immunotherapies, convalescent plasma treatment and treatment discontinuation will be ignored (treatment policy strategy).

Patients with unknown death status at Day 60 visit will be treated as alive for the analysis if recovered during the study. Specifically, all patients who recovered, were discharged, and/or had 30 day follow up where they remained recovered, will be imputed at Day 60 as being alive/recovered if data is missing. Patients who discontinued from study before or on Day 1 and Day 60 survival follow up status as unknown are excluded from the primary analysis. This is not expected to introduce a bias in favor of either treatment arm since the clinical status is not likely to be affected within two days by the study treatment administered on Day 1.

To protect the trial level type 1 error rate at α=5% (two sided) between the two tests of the primary endpoint analysis and the key secondary endpoint analysis, the Benjamini and Hochberg testing strategy is used. Test statistics for Time to recovery and All-cause mortality at Day 60 are positive correlated and Hochberg testing strategy will protect the trial level type 1 alpha level.

8.5.3 Sensitivity Analyses

There are three planned sensitivity analyses that both target the primary estimand. The first will be to relax the assumption that the baseline imputed PaO_2/FiO_2 is a confounding factor; the analysis will therefore be repeated but without stratification by the baseline imputed PaO_2/FiO_2 of ≤100 vs. >100. This analysis implies the baseline risk is the same across strata. The second sensitivity analysis is to examine the potential effects of missing death data due to dropout using tipping point sensitivity analysis. In the third analysis, day 60 death rate for each treatment group will be estimated by the Kaplan-Meier procedure. Hypothesis testing will be based on the Kaplan-Meier estimates and standard errors estimated by Greenwood formula using the log-log transformation of the survival function stratified by the baseline imputed PaO_2/FiO_2 of ≤100 vs. >100.

8.5.4 Supplementary Analyses

There are two planned supplementary analyses.

The first supplementary analysis changes the population to the safety analysis set. The analysis will be the same analysis as the main analysis except that treatment group will be based on actual treatment received.

The second supplementary analysis changes the population to the efficacy analysis set including 26 subjects from baseline imputed PaO_2/FiO_2 of ≥ 201 stratification group and the same analysis as the main analysis will be performed.

8.5.5 Supportive Secondary Efficacy Analysis

8.5.5.1 All-cause mortality at Day 30 visit

The analysis will be the same as the Day 60 visit analysis.

8.5.5.2 Proportion of patients requiring invasive mechanical ventilation or dying during the 60 days from the SFISD

Time to requiring invasive mechanical ventilation is defined as time from SFISD to the first time the patient requiring invasive mechanical ventilation or dying. Proportion of patients requiring invasive mechanical ventilation or dying is defined as the proportion at study Day 60. Day 60 rate for each treatment group will be estimated by the Kaplan-Meier procedure. Hypothesis testing will be based on the Kaplan-Meier estimates and standard errors estimated by Greenwood formula using the log-log transformation of the survival function stratified by the baseline imputed PaO_2/FiO_2 of ≤ 100 vs. >100. Subjects who do not have death or require invasive mechanical ventilation in the record will be censored on the last known status date (administrative censoring). Patients who recovered, were discharged, had 30 day follow up where they remained recovered, and did not use invasive mechanical ventilation during the study, will be censored at Day 65 as being alive/recovered without event if data is missing.

8.5.5.3 Proportion of patients requiring invasive mechanical ventilation during the 60 days from the SFISD

The analysis will be the same as the proportion of patients requiring invasive mechanical ventilation or dying during the 60 days from the SFISD. Subjects who have death and do not require invasive mechanical ventilation will be censored on the date of death (hypothetical strategy). Patients who recovered, were discharged, had 30 day follow up where they remained recovered, and did not use invasive mechanical ventilation during the study, will be censored at Day 65 as being alive/recovered without event if data is missing.

8.5.5.4 Differences in Outcomes Measured by an 8-Point Ordinal Scale at Day 12

Proportion of patients in each category of 8-point ordinal scale at Day 12 will be calculated for each treatment group, and the odds ratio will be estimated. The proportion will be compared between the 2 treatment groups using a proportional odds model with fixed factor of treatment groups.

8-point ordinal scale will be imputed based on LOCF method based on the scale at discharge if the patient is discharged from hospital with the ordinal scale of 7 or 8 (composite strategy). 8-point ordinal scale will be imputed as 1 after the death (composite strategy).

8.5.5.5 Differences in Outcomes Measured by an 8-Point Ordinal Scale at Day 30

The analysis will be the same as the Day 12 analysis but extended at Day 30.

8.5.5.6 Number of Days in the Hospital

The date of each hospital admission and discharge will be collected for each patient. The number of days when patients are in hospital during the first 28 Days of the study will be summarized by treatment group and compared between the 2 treatment groups using an analysis of variance model, which includes treatment group in the model, which includes treatment group as the fixed effect in the model.

Two sets of analyses will be conducted. The first analysis is to assess patient benefit; the number of days in the hospital will be defined as 28 if the patient dies due to COVID-19 (composite strategy). The second analysis is to assess healthcare systems benefit; the number of days in the hospital before the patient's death will be used in the analysis (while alive strategy).

8.5.5.7 Number of Days in the Hospital

The analysis will be the same as the number of days in the hospital

8.6 Tertiary Endpoint Analysis

8.6.1 Change in Imputed PaO_2/FiO_2 Ratio

SpO_2 and FiO_2 will be examined at each scheduled time point. Worst value within the previous 24 hours at each scheduled time point will also be recorded. Conversion from SpO_2 to PaO_2 will follow Table 3.

Table 3 Look up table for conversion SpO_2 to PaO_2

Measured SpO_2 (%)	Imputed PaO_2 (mmHg)
100*	167*§
99*	132*
98*	104*

97*	91*
96	82
95	76
94	71
93	67
92	64
91	61
90	59
89	57
88	55
87	53
86	51
85	50
84	49
83	47
82	46
81	45
80	44
79	43
78	42
77	42
76	41
75	40
74	39
73	39
72	38
71	37
70	37
69	37

* Generally considered unreliable on the basis of the sigmoidal shape of the hemoglobin-oxygen dissociation curve
§ Based on SpO_2 99.5%.

Mean changes from baseline to each post-baseline visit and time point will be summarized by treatment groups for both at visit imputed PaO_2/FiO_2 value and worst value within the previous 24 hours of the visit imputed PaO_2/FiO_2 value.

Change from baseline over time in the imputed PaO_2/FiO_2 will be analyzed using a mixed effect model repeat measurement (MMRM) model with an unstructured covariance structure (if model is estimable). The analysis model will include Baseline value of the endpoint as a covariate, study day and treatment as fixed effect, study day by treatment as interaction, and subjects as random effect. Point estimates, 95% CIs and p-values for the difference between the Auxora treated patients and local standard care group at each scheduled time point will be obtained.

Similar analysis to the above analysis will be conducted for the change from baseline over time in the imputed PaO_2/FiO_2 endpoint using the last observation carried forward method as the missing value imputation method.

8.7 Safety Analysis

The safety evaluation will be purely descriptive using descriptive statistics (N, mean, SD, median minimum and maximum) or frequency tables where appropriate using Safety Analysis Set, unless otherwise specified. No imputation will be made in case of missing values.

8.7.1 Study Medication Exposure

Number of doses of double-blind study medication will be summarized in table and presented in data listing for the safety population.

8.7.2 Adverse Events

Adverse events will be coded using MedDRA dictionary. If an incomplete or missing onset date was collected for an AE, the imputation method of missing data specified in Section 8.1.4.4 will be applied. Treatment-emergent AEs (TEAEs) are defined as AEs that started or worsened in severity on or after the SFISD.

The relationship to study treatment for each AE will be classified as 'Unrelated', 'Unlikely', 'Possible', 'Probable', 'Definite'. AEs classified as 'Possible,' 'Probable' and Definite will be analyzed as 'Related' in the AE summaries. Data listings, patient narratives, etc. will present the relationship to study treatment as collected on the CRF.

8.7.2.1 Tabulations of TEAEs

The frequency and incidence of TEAEs will be summarized by System Organ Class (SOC) and Preferred Term (PT) by treatment and then by overall treatment. The actual version of the MedDRA coding dictionary used will be noted in the clinical study report. The SOC is sorted by alphabetic order; then within SOC, PT is sorted by descending counts under Auxora column, then descending counts under Placebo column, then alphabetic order for PTs with the same count.

Each of the summaries will be done at the subject level. Multiple occurrences of the same event within a subject will be counted once in the summaries by System Organ Class and PT.

All AEs for each subject, including multiple occurrences of the same event, will be presented in full in a comprehensive listing including subject number, treatment, severity, seriousness, action taken, outcome, relationship to treatment, onset/stop date and duration.

The data listings for serious TEAE and TEAE leading to discontinuation of study treatment will be generated as well.

8.7.2.2 Adverse Event Overview

An overview of AEs will be presented by the actual treatment received overall and consisting of the number and percentage of patients experiencing at least one event for the following AE categories:

- Any TEAEs
- TEAEs related to study treatment
- Serious TEAEs
- Serious TEAEs related to study treatment
- TEAEs leading to discontinuation of study treatment
- TEAEs leading to death

8.7.2.3 Adverse Event by SOC and PT

The following summaries of AEs will be generated:

- Incidence of TEAEs
- Incidence of TEAEs related to study treatment
- Incidence of serious TEAEs
- Incidence of serious TEAEs related to study treatment
- Incidence of TEAEs leading to discontinuation of study treatment
- Incidence of TEAEs leading to death
- Incidence of severe TEAEs

8.7.3 Additional Safety Assessments

8.7.3.1 Vital Signs

Vital signs will include temperature, systolic and diastolic blood pressure, respiratory rate (breaths per minute), heart rate (beats per minute).

Vital signs will be examined at each scheduled visit and time point. Clinically-significant, treatment-emergent findings will be reported as AEs.

In addition, vital signs results will be flagged as potentially clinically significant (PCS) if they meet the criteria which are defined below.

- Pulse rate >140 bpm.
- Pulse rate <45 bpm.

- Diastolic blood pressure >100 mmHg.
- Systolic blood pressure >155 mmHg.
- Systolic blood pressure <60 mmHg.
- Body temperature ≥39°C.

The number and percent of patients meeting each PCS criterion at each scheduled visit and time point will be summarized by the actual treatment and then by the overall treatment.

8.7.3.2 Laboratory Analyses

CBC with Differential, Serum Chemistries, Ferritin, C-reactive protein, Procalcitonin and D-Dimer will be performed.

Laboratory values by visit will be summarized as boxplots. The lower and upper reference range values for lab tests will be used to grade the lab results to low, normal and high categories. Shift tables will display:

1. Shift from baseline category to the worst category;
2. Shift from baseline category to the last category.

8.8 Other Analyses

8.8.1 Other Variables and/or Parameters

Biomarkers will be measured at baseline and with subsequent assessments at various timepoints up to 240 hours. Log-transformation will be applied when necessary. Change from baseline over time will be analyzed using a mixed effect model repeat measurement (MMRM) model with an unstructured covariance structure (if model is estimable). The analysis model will include Baseline value of the endpoint as a covariate, study day and treatment as fixed effect, study day by treatment as interaction, and subjects as random effect. Point estimates, 95% CIs and p-values for the difference between the Auxora treated patients and local standard care group at each scheduled time points will be obtained.

8.8.2 Subgroup Analyses

Subgroup analyses will be performed to explore how time to recovery through Day 60 visit and mortality at Day 60 visit are influenced by baseline variables and to evaluate the treatment effect at different levels of each of these variables. The Kaplan-Meier analysis, Cox model and CMH test will be performed by subgroup levels of the baseline variables listed below:

- Age (<65, ≥65)
- Baseline CRF calculated imputed PaO_2/FiO_2 (≤100, 101-200, ≥201)

- Race (White, Black, Asian, Other).
- Gender (Male, Female)
- BMI (<30, ≥30)
- Remdesivir (being already administered at randomization, administration started after randomization, never administered)
- Dexamethasone (being already administered at randomization, administration started after randomization, never administered)
- Remdesivir and Dexamethasone (being already administered together at randomization or administered together starting after randomization)
- Convalescent plasma (being administered at randomization, administration started after randomization, never administered)

9 INTERIM ANALYSIS

9.1 Interim Sample Size Re-estimation and Futility Analysis

One unblinded interim sample size re-estimation is planned and was conducted by the IDMC when 70 patients in the baseline imputed PaO_2/FiO_2 ≤200 subgroup have completed the study.

In the protocol plan, the conduction power of the time to recovery will be calculated based on the conditional power analysis of the time to recovery by assuming the interim estimated parameter values as the estimates of these parameters. Sample size will be re-estimated at the interim analysis to provide a conditional power of 90%, based on the interim evaluation of the treatment efficacy. The study will use a sequential design with O'Brien-Fleming critical boundary (O'Brien and Fleming 1979) at the interim analysis to control type I error. In case of sample size modification, the final critical boundary will be adjusted to control type I error (Gao et al., 2008). Based on the estimated standard error of 0.25 from Part 1 data, the relationship between the observed recovery ratio at interim analysis and the re-estimated sample size to reach 90% conditional power is shown in the following table:

Table 4 Relationship Between the Observed Recovery Ratio at Interim Analysis and the Re-estimated Sample Size to Reach 90% Conditional Power

O'Brien-Fleming boundary at the interim analysis, interim analysis at information fractions 0.25, Standard error of log(recovery ratio) =0.25				
Observed recovery ratio at interim analysis	Conditional power based on interim estimation	Sample size (events) to reach 90% conditional power based on interim estimation	Conditional power based on recovery ratio = 1.37	Sample size (events) to reach 90% conditional power on recovery ratio = 1.37
1.57	97%	214 (178)		
1.47	90%	290 (241)		

1.44	86%	320 (266)		
1.40	80%	384 (318)		
1.39	78%	400 (332)		
1.37	67%	438 (363)		
1.34	59%	500(415)		
1.31	59%	600(503)		
1.20	28%	1414(1174)	64%	507(421)
1.10	8%	5648(4688)	56%	563(468)
1.05			52.2%	596(495)
1.00			48%	630(523)

Based on the above relationship, the following sample size re-estimation rule is proposed for this study at the interim analysis of the 70 patients in the imputed PaO_2/FiO_2 ≤200 subgroup:

- Recovery rate ratio <1.05, increase sample size >600 for imputed PaO_2/FiO_2 ≤200 subgroup
- Recovery rate ratio 1.05 to <1.34, increase sample size to 600 for imputed PaO_2/FiO_2 ≤200 subgroup
- Recovery rate ratio 1.34 to 1.38, increase sample size to 500 for imputed PaO_2/FiO_2 ≤200 subgroup
- Recovery rate ratio 1.39 to 1.43, increase sample size to 400 for imputed PaO_2/FiO_2 ≤200 subgroup
- Recovery rate ratio ≥1.44, do not increase sample size for imputed PaO_2/FiO_2 ≤200 subgroup

The sample size re-estimation rule was updated and a non-binding futility analysis was added prior to the conduct of the interim sample size re-estimation. The updated sample size re-estimation rule was based on the promising zone with re-estimated sample size of 500 patients. The study used a sequential design with O'Brien-Fleming critical boundary (O'Brien and Fleming 1979) at the interim analysis to control type I error. In case of sample size modification, the final critical boundary will be adjusted to control type I error (Gao et al., 2008). The detailed promising zone rule is described in the separate interim statistical analysis plan and was finalized before the interim sample size re-estimation.

9.1.1 Critical Boundary Adjustment

If the sample size is modified to a new information time T_{new}, the second critical boundary, $c_2 = 1.96$ will be modified according to Gao et al., 2008:

$$C_{new} = \frac{1}{\sqrt{T_{new}}}\left\{\frac{\sqrt{T_{new}-t_{interim}}}{\sqrt{T-t_{interim}}}\left(C_2\sqrt{T} - s_{interim}\right)\right\} + \frac{s_{interim}}{\sqrt{T_{new}}}.$$

This adjustment properly controls the type I error.

10 VALIDATION

The tables, figures and listings (TFLs) planned in this SAP will be programmed using SAS software version 9.4 (or above). The TFLs will be quality checked by the statistics team at Princeton Pharmatech using SAS software version 9.4 (or above).

11 PROGRAMMING AND DATA PRESENTATION CONVENTIONS

Listings will be presented in treatment, subject, visit (where applicable) and date (where applicable) order. Listings will be produced (landscape in MS Word) using PROC REPORT.

Summary tabulations will be presented by treatment group (and overall if appropriate), scheduled visit order (if appropriate). Continuous data summaries will present (unless stated otherwise) number of observations, mean, standard deviation, median, minimum and maximum. Categorical data summaries will present the number of observations and the corresponding percentage.

12 BIBLIOGRAPHY

Gao P, Ware JH, Mehta C. Sample size re-estimation for adaptive sequential design in clinical trials. Journal of biopharmaceutical statistics. 2008;18(6):1184-96.
doi: 10.1080/10543400802369053.

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549-56.