CM4620-203 Statistical Analysis Plan

A Randomized, Double-Blind, Placebo Controlled Dose Ranging Study of Auxora in Patients with Acute Pancreatitis and Accompanying Systemic Inflammatory Response Syndrome (CARPO)

Version: 1.0

Date: June 10, 2024

Clinical Trial Number: NCT04681066

Statistical Analysis Plan

Protocol CM4620-203

A Randomized Double-Blind, Placebo Controlled Dose-Ranging Study of Auxora in Patients with Acute Pancreatitis and Accompanying Systemic Inflammatory Response Syndrome (CARPO)

PREPARED BY:

NAME: **Jeffrey Zhang, PhD.**

POSITION: **Principal Statistician, Princeton Pharmatech**

VERSION: **Final 1.0**

DATE: **10 June 2024**

APPROVAL FORM

STATISTICAL ANALYSIS PLAN

Protocol No. CM4620-203

Protocol Title **A Randomized Double-Blind, Placebo Controlled Dose-Ranging Study of Auxora in Patients with Acute Pancreatitis and Accompanying Systemic Inflammatory Response Syndrome (CARPO)**

Version Final 1.0

Date 10 June 2024

Author Jeffrey Zhang, PhD, Principal Statistician, Princeton Pharmatech

DocuSigned by:
Jeffrey Zhang, PhD
Signer Name: Jeffrey Zhang, PhD
Signing Reason: I approve this document
Signing Time: 6/10/2024 | 8:14:48 AM PDT
17A8AB3000DE4D7C97C78888DF573F10

6/10/2024

Signature Date

APPROVED BY:

DocuSigned by:
Sudarshan Hebbar
Signer Name: Sudarshan Hebbar
Signing Reason: I approve this document
Signing Time: 6/10/2024 | 8:10:56 AM PDT
732B16E369D74416B6B6AE164D8A9C32

6/10/2024

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

TABLE OF CONTENTS

TABLE OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
AP	Acute pancreatitis
ARDS	Acute respiratory distress syndrome
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
CRO	Contract Research Organization
CS	Clinically significant
CSR	Clinical study report
DP	Decimal places
eCRF	Electronic case report form
ECMO	Extracorporeal membrane oxygenation
IDMC	Independent data monitoring committee
IWRS	Interactive web response system
LAR	Legally authorized representatives
LOCF	Last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
PCS	Potentially clinically significant
PK	Pharmacokinetic(s)
PT	Preferred term
QC	Quality control
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 CM4620-203 Protocol version 4 dated 05 October 2023. The plan covers statistical analysis, tabulations and listings of safety and efficacy data to assess the safety and efficacy of Auxora compared to placebo in patients with acute pancreatitis and accompanying systemic inflammatory response syndrome (CARPO).

The statistical analyses and production of the outputs described in the SAP will be conducted and quality control (QC) reviewed, using SAS version 9.4 or higher.

2. STUDY OBJECTIVES

2.1 Primary Objectives:

- To assess the dose response and efficacy of three different dose levels of Auxora in patients with acute pancreatitis and accompanying SIRS;
- To assess the time to medically indicated discharge in patients who are responders to early tolerance of solid food intake versus non-responders.

2.2 Secondary Objective:

- To assess the safety and tolerability of varying doses of Auxora in patients with acute pancreatitis and accompanying SIRS

3. STUDY DESIGN

3.1 Overall Study Design

This double blind, randomized, placebo-controlled study will evaluate the efficacy, safety, and tolerability of three different dose levels of Auxora in patients with acute pancreatitis and accompanying SIRS.

Approximately 216 patients will be randomized 1:1:1:1 into one of 4 groups using a computer generated randomization scheme accessed through an interactive voice/web response system (IXRS). Randomization will be first stratified by sex (male or female) and then by risk for organ failure in the sex subgroups (higher or lower). Higher risk for organ failure is defined by the presence of both an elevated hematocrit (HCT ≥44% for men or ≥40% for women) and hypoxemia (imputed PaO_2/FiO_2 ≤360). Lower risk for organ failure is defined by the absence of either or both an elevated hematocrit and hypoxemia. The PaO_2/FiO_2 will be determined using an arterial blood gas or imputed using pulse oximetry.

All patients will have received a Screening CECT/MRI of the abdomen/pancreas before being randomized into the study. CECTs/MRIs performed as standard of care may be used as the Screening CECT/MRI but must have been performed in the 24 hours before Consent or after Consent and before Randomization.

The Start of First Infusion of Study Drug (SFISD) should occur within 8 hours of the patient or LAR providing informed consent. Patients randomized to Group 1 will receive 2.0 mg/kg (1.25 mL/kg) of Auxora intravenously every 24 hours (±1 hour) for a total of three doses. Patients

randomized to Group 2 will receive 1.0 mg/kg (0.625 mL/kg) of Auxora intravenously every 24 hours (±1 hour) for a total of three doses. Patients randomized to Group 3 will receive 0.5 mg/kg (o.3125 mL/kg) of Auxora intravenously every 24 hours (±1 hour) for a total of three doses. Patients randomized to Group 4 will receive emulsion without any active pharmaceutical ingredient (placebo). Patients in Group 4 will receive one of three randomly assigned dose volumes, 1.25 mL/kg, 0.625 mL/kg, or 0.3125 mL/kg, which will be administered intravenously every 24 hours (±1 hour) for a total of three doses. 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, the upper limit of the volume of Auxora and volume of placebo that will be administered will be 156.25 mL. The sponsor, investigators and patients will be blinded to the assigned group. In the event of a medical emergency, investigators will be able to receive the treatment assignment if required to provide optimal care of the patient.

For all 4 groups, a study physician or appropriately trained delegate will perform assessments at screening, at the baseline assessment, immediately prior to the SFISD, and then every 24 hours until 240 hours after the SFISD, or until discharge if earlier. If patients remain hospitalized at Day 12, assessments will then be performed every 48 hours starting on Day 12 until Day 28, or until discharge if earlier. Patients discharged from the hospital before Day 25 will return at Day 30 (+5 days) to perform the Day 30 assessments. If patients are discharged on Days 25-29, the Day 30 assessments may be performed prior to discharge.

Patients will receive another CECT of the abdomen/pancreas at the Day 30 (±5 days) visit. All CECTs performed as standard of care after randomization and before the Day 30 CECT will also be captured. A blinded central reader will read the Screening, Day 30, and any standard of care CECTs obtained between randomization and 30 days.

Patients will complete the modified American Neurogastroenterology and Motility Society Gastrointestinal Cardinal Symptom Index Daily Diary (mGCSI-DD) worksheet at the baseline assessment, at 96 hours, 168 hours, Day 14 and Day 21 (for patients who remain hospitalized on these days), on the day of discharge, and daily at bedtime after discharge until the Day 30 visit. Patients who are discharged on Days 25-29 will not complete the mGCSI worksheet after discharge.

It is recommended that all patients randomized in the study should receive care consistent with the 2018 American Gastroenterological Association (AGA) Institute Technical Review of the Initial Medical Management of Acute Pancreatitis. Patients should receive local standard of care (SOC) for the management of other medical conditions.

In patients with acute pancreatitis, the AGA strongly recommends early oral feeding (within 24 hours) rather than keeping the patient nil per mouth (Nil per Os, NPO). Patients randomized into the study, therefore, will be offered a low fat, ≥500-calorie solid meal at each mealtime after the infusion of the first dose of study drug if alert and not on mechanical ventilation. If the patient does not wish to eat the solid meal when offered or is unable to tolerate the solid meal, they should then be offered a liquid meal. The same approach should occur at each subsequent mealtime. When patients eat a solid meal, it should be recorded if they ate ≥50% of the meal and

if they either vomited or experienced an increase in abdominal pain in the two hours after eating a meal.

It is also recommended that all patients randomized in the study should not be discharged from the hospital until solid food is tolerated, abdominal pain has resolved or been adequately controlled, and there is no clinical evidence of infection. Tolerating solid food is defined as eating ≥50% of a low fat, ≥500-calorie solid meal without an increase in abdominal pain or vomiting. If the patient is not tolerating either solid or liquid meals, tube feedings should be considered.

All protocol required laboratory testing, except biomarker and PK samples, will be performed at the local laboratory. Results from the biomarkers and PK blood samples collected as part of the protocol and being tested at a central lab will not be available to assist the PI or treating physician in managing the patient.

3.2 Schedule of Study Procedures

Table 1. Schedule of Assessments

	Screening	If Eligible, Baseline Assessment	R	SFISD	24 (±1) Hours[i]	48 (±1) Hours[i]	72 (±2) Hours	96, 144, 192, 240 (±4) Hours[h,k]	120, 168, 216 (±4) Hours[h,k]	Days 12-28 Q48 (±4) Hours[h,k]	Day 30 (±5) Days
Informed Consent	X										
Demographics	X										X[l]
Medical History	X										
List of Medications	X										
Concomitant Meds				X	X	X	X	X	X		X
Height		X									
Weight		X					X				X
Vital Signs	X				X	X	X	X	X		X
Abdominal Exam	X										
PNRS		X			X	X	X	X	X		X
Serum Lipase	X[a]										
CBC+diff+Platelets	X[a]				X[a]	X[a]	X[a]		X[a]		X[a]
Presence of SIRS[b]	X										
SpO_2+FiO_2		X			X	X	X	X	X		X
CECT of Pancreas[c]	X										X
Serum Pregnancy Test	X[a]										
Randomize Patient[d]			X								
Serum Chemistries		X[a]			X[a]	X[a]	X[a]		X[a]		X[a]
Serum Procalcitonin		X[a]			X[a]	X[a]	X[a]		X[a]		X[a]
Serum Biomarker[e]		X			X	X	X		X		X
Urine Biomarker[e]		X			X	X	X		X		X
Administer Study Drug				X	X[i]	X[i]					

PK sample[e]				X[j]	X[j]		X[j]				X[j]
Medically Indicated Discharge Criteria[f]					X	X	X	X	X	X	X
Enhanced Recovery Strategy[g]				X	X	X	X	X	X	X	X[m]
mGCSI-DD[h]		X[h]			X[h]					X[h]	X[h]
AE/SAE Evaluation			X	X	X	X	X	X	X	X	X
Telephone Contacts, as needed		Following patient discharge, Investigator or designee should contact patient at least once weekly to confirm diary completion, Day 30 visit schedule and assess for AE/SAEs.									

AE = adverse event; CECT = contrast-enhanced computed tomography; PK = pharmacokinetics; mGCSI-DD = Modified American Neurogastroenterology and Motility Society Gastrointestinal Cardinal Symptom Index Daily Diary; PI = Principal Investigator; PNRS = Pain Numeric Rating Scale; SAE = serious adverse event; SFISD = Start of First Infusion of Study Drug

a. Blood sample for serum pregnancy test, serum lipase and CBC+diff+Platelets at Screening will be analyzed at local laboratory. May use results obtained as standard of care if blood drawn in the previous 12 hours prior to consent or following consent but prior to randomization. Following screening, all protocol-required blood samples, except as noted in Footnote "e", will be analyzed at local laboratory. Baseline blood samples including biomarkers may be performed on blood drawn in the previous 12 hours of randomization. For all later assessment times, blood drawn in the previous 12 hours maybe used (except PK).
b. The most extreme value for each criterion in the 24 hours before consent or following consent but prior to randomization may be used.
c. A CECT already performed in the 24 hours prior to the patient or LAR providing informed consent may be used as the Screening CECT. In addition to the Screening and the Day 30 visit CECTs, CECTs performed as standard of care after randomization and to 25 days after SFISD will be captured and submitted for central reading. If the CECT is performed 25 to 29 days after SFISD, it will substitute for the Day 30 visit CECT scan and will be captured.
d. Hematocrit collected as standard of care within 12 hours of consent or following consent but prior to randomization may be used for Inclusion Criterion 3c.
e. Biomarker blood and urine samples and blood samples for PK will be analyzed at central laboratory. If site or patient does not participate in these assessments this will be documented at start of study and will not be considered a protocol deviation.
f. Determine if patient had met the medically indicated discharge criteria within the prior 24- or 48-hour visit assessment.
g. Patients will be offered a low fat, ≥500-calorie solid meal at each mealtime after the infusion of the first dose of study drug until discharge if alert and not on mechanical ventilation, or if not NPO for a planned surgery/medical procedure, or if not NPO because of an a cute medical condition. If the patient does not wish to eat the solid meal when offered or is unable to tolerate the solid meal, they should then be offered a liquid meal. Determine for each meal if the patient ate ≥50% of the meal without vomiting or having abdominal pain in the two hours after the meal.
h. The mGCSI-DD score will be completed by all eligible patients at the baseline assessment and on the day of discharge, which will vary by patient as shown in schedule above. In addition, if the patient remains hospitalized at 96 and/or 168 hours and/or at Day 14 and/or Day 21 patient should be requested to record their symptoms at bedtime using the mGCSI-DD worksheet. Patients will be trained on completing the mGCSI-DD worksheet. Upon discharge, patient will record the mGCSI-DD score daily before bedtime until the Day 30 visit.
i. In the scenario where the PI or treating physician wants the infusion(s) to begin earlier than the 24 +/- 1 hour window, the PI or treating physician should contact the medical monitor to discuss whether it is appropriate to begin infusion earlier (up to 6 hours earlier) than the 24 +/- 1 hour protocol window.
j. Blood samples for PK analysis will be drawn as soon as possible after the completion of the first infusion of study drug (within +2 hours), prior to the start of second infusion of study drug (within -30 min), at 72 hours (±2 hours) from the SFISD, and at the Day 30 visit. If the patient is discharged after the third infusion of study drug and prior to 72 hours from the SFISD, the blood sample will be drawn prior to discharge.
k. Study assessments continue daily, or every other day, as noted in schedule of events until discharge or the Day 30 visit. All patients return for Day 30 assessments if discharged earlier than Day 25.
l. At the Day 30 visit (or at the time of death if before Day 30), record the Investigator's opinion on the etiology of the episode of AP.
m. Only patients who remain hospitalized at Day 30.

4. SAMPLE SIZE

A sample size of 216 patients will be randomized into four groups on a 1:1:1:1 basis, resulting in 54 patients randomized into each group. This sample size will provide 86% power for testing the difference in two populations having a median length of time of 72 and 144 hours, respectively, to tolerating solid food. This sample size will provide 80% power with a two-sided alpha of 0.05 to detect a 45% response rate for tolerating solid food in an Auxora dose group versus a 20% response rate in the placebo group in the 72 hours after completion of the first study drug infusion.

5. STUDY ENDPOINTS

5.1 Efficacy Endpoints:

- Primary endpoint
  - Time to solid food tolerance
- Secondary endpoints:
  - Solid food tolerance at 48 hours, 72 hours, and 96 hours after the Start of First Infusion of Study Drug (SFISD) and at discharge
  - Time to medically indicated discharge
  - Length of stay in the hospital
  - Length of stay in the intensive care unit (ICU) for patients admitted to the ICU
  - Re-hospitalization for acute pancreatitis (AP) by Day 30
  - Change in severity of AP by CTSI score from screening to Day 30
  - Development of pancreatic necrosis ≥30% and >50%
  - The persistence of SIRS ≥48 hours after the SFISD
  - Incidence, severity, and duration of organ failure
  - Mortality by Day 30
  - Change in pain score and opioid use
- Exploratory endpoints:
  - Development of infected pancreatic necrosis
  - Development of sepsis
  - Hospital procedures for the management of pancreatic necrosis
  - Change in GCSI-DD score
  - Change in albumin
  - Change in absolute neutrophil count (ALC)/absolute lymphocyte count (ALC) ratio
  - Change in IL-6 levels
  - Change in urine neutrophil gelatinase-associated lipocalin (NGAL) levels

5.2 Safety Endpoints:

- The incidence of TEAEs and SAEs
- The intensity and relationship of TEAEs and SAEs
- Clinically significant changes in vital signs and safety laboratory results

5.3 Proof of Concept Composite Efficacy Endpoint:

The proof-of-concept efficacy analysis uses a hierarchical combination applying the method of Finkelstein-Schoenfeld (Finkelstein 1999) to:

- All-cause mortality at Day 30 Visit, and
- Persistent severe respiratory failure through Day 30 Visit, and
- Pancreatic necrosis ≥30% through Day 30 Visit, and
- Rehospitalization for AP through Day 30 Visit, and
- Time to solid food tolerance within 30 Days of the study

6. ANALYSIS POPULATION(S)

6.1 Efficacy and Safety

The efficacy analyses will be performed on the Modified Intent-to-treat Population (MITT) consisting of all randomized patients who receive any amount of study drug. The safety analyses will be performed on the MITT data set.

7. TIMINGS OF ANALYSES

7.1 Final Analysis

The final analysis will be conducted when all randomized patients complete the study.

8. STATISTICAL ANALYSIS

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), standard error (SE), minimum and maximum for continuous data, interquartile ranges for nonparametric 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% unless specified otherwise. 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 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 time of the SFISD. If multiple measurements are recorded on the same day, the last measurement recorded prior to or at the time of the SFISD 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.

8.1.4.2 Study Time

For patients enrolled in the study, the first dose of study medication will be administered at 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 database. 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 for Safety Endpoints and Concomitant Medications

Missing data will not be imputed in the summary of safety endpoints at each time point. 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 Study Visits

The study summary and analysis will be based on the eCRF visits.

8.1.7 Pooling of Sites

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

8.2 Subject Disposition and Withdraws

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 discontinuation reasons in tables for each treatment group and then total subjects. The data listing for subject discontinuation will be generated.

8.3 Demographic and Baseline Characteristics

Patient demographic and baseline characteristics will be summarized by mean, SD, median, minimum, and maximum for continuous variables; interquartile ranges for nonparametric data, 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, the number of SIRS criteria met at Screening: two versus greater than two, number of patients at high risk for organ failure [defined by the presence of both an elevated hematocrit (HCT ≥44% for men or ≥40% for women) and hypoxemia (imputed PaO_2/FiO_2 ≤360)], number of patients with baseline hematocrit value and category (<44%, >=44% for males; <40%, >=40% for females), the number of patients with abdominal guarding or rebound tenderness, and the number of patients

with CECT at Screening with peripancreatic fluid or pleural effusion by central reading. The Modified Intent to Treat (MITT) Analysis Set will be used for the summaries.

Other baseline variables such as relevant medical history or other medically relevant criteria, such as the cause of acute pancreatitis, will also be included in the baseline tables.

8.4 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 selected concomitant medications will be summarized overall and by Anatomic Therapeutic Chemical (ATC) Classification level 2 terms, preferred drug names and treatment groups. A separate listing for selected concomitant medications will also be provided.

8.5 Study Medication Exposure

The number of treatment doses, delayed doses, reduced doses, and interrupted doses per patient of double-blind study medication will be summarized in the table for the MITT population. The percentage of delayed doses, reduced doses, and interrupted doses per patient and per total number doses will be summarized.

8.6 Primary Endpoint Analysis

8.6.1 Generalized Multiple Comparisons and Modelling (gMCP-Mod)

The primary objective of this study is to demonstrate a dose-response effect of Auxora compared to placebo of the primary endpoint time to solid food tolerance through Day 30. To demonstrate the dose-response relationship between treatment with Auxora compared to placebo for the primary endpoint, the primary analysis will be based on a generalized Multiple Comparisons and Modeling (gMCP-Mod) analysis [Bretz et al., 2005; Pinheiro et al., 2014]. This analysis tests for a dose-response relationship, allowing for uncertainty in the dose-response relationship through inclusion of contrasts from multiple pre-specified candidate models assessing dose-response.

The gMCP-Mod procedure extends the MCP-Mod methodology to non-continuous endpoints. This is accomplished by decoupling the dose-response model from the expected response through the use of an ANOVA-style parameterization of the dose-response parameter. For the time-to-event endpoint used in this study, a stratified Cox regression of the time to solid food tolerance at each dose level will be used, stratified by the CRF recorded sex (male or female) and then by risk for organ failure based on CRF data in the sex subgroups (higher or lower). After this first step, the mean estimates of log hazard ratio from the Cox regression model and the corresponding covariance matrix, both of which are on the log scale, will then be used during the rest of the gMCP-Mod procedure. Point estimates from the selected models will be displayed on the ratio scale in all statistical outputs.

The gMCP-Mod (Pinheiro et al., 2014) will be employed in a three-step approach in the manner described below:

- Step 1: Conventional Step – Hazard Ratio Between Each Dose Versus Placebo

  The data are analyzed using stratified Cox regression to obtain treatment effects and corresponding variance covariance matrix. The Cox regression will be stratified by sex (male or female) and the risk for organ failure (higher or lower).

  The estimated treatment hazard ratio and the associated 95% confidence intervals will be presented for the treatment contrast of each dose of Auxora versus placebo

- Step 2: Proof-of-concept Step – Candidate Models and Multiple Testing Procedure

  The null hypothesis of flat dose-response relationship as compared to placebo for the primary efficacy endpoint will be tested at a significance level of one-sided 15% against the alternative hypothesis of a non-constant dose-response curve using a multiple contrast test as described in the gMCP-Mod methodology (Pinheiro et al., 2014). The contrast test statistic is a linear combination of the optimal contrast coefficients corresponding to the candidate dose-response curve with the hazard ratio at each individual dose obtained from Step 1 (conventional step).

  The contrast coefficients are chosen to maximize the power to detect pre-specified candidate models. There are two candidate models that will be considered for the primary analysis with the optimal contrasts shown in following table. These optimal contrasts will be updated using the expected model means from the candidate set of the two models and the estimated variance-covariance matrix from the collected data. The updated contrast coefficients will be described in the final CSR.

  The optimal contrast based on Candidate models for Auxora dose for log(Hazard Ratio):

Model	Contrast Coefficients for Dose		
Dose	0.5 mg/kg	1.0 mg/kg	2.0 mg/kg
Emax	0.508	0.588	0.629
sigEmax	-0.611	0.488	0.623

  Note: Candidate model of Emax model is based on the assumption of parameter of ED50=0.026 mg/kg. Candidate model of sigEmax mode is based on the assumption of parameters of ED50=0.707 mg/kg and h=6.34.

  The proof of concept (PoC) is established if at least one dose-response model is statistically significant, thereby rejecting the null hypothesis of a flat dose-response curve and indicating a benefit of Auxora over placebo.

  The global test decision is based on the maximum of the contrast test statistics. A critical value $\boldsymbol{q}$ controlling the type I error rate can be derived from the fact that the contrast test statistics approximately follow a multivariate normal distribution and that their maximum

follows the distribution of the maximum of a multivariate normal distribution. If the maximum contrast test statistic exceeds the critical value q, the overall null hypothesis of a flat dose-response curve is rejected, and one proceeds to characterize the dose-response efficacy relationship.

For each candidate dose-response curve, the test statistics and corresponding adjusted p-values will be presented.

Once proof of concept has been established, then subsequently, it will be tested whether at least one dose of Auxora achieves time to solid food tolerance superior to placebo (i.e., the maximum anticipated effect of Auxora for achieving the primary endpoint, based on the best fitting monotonic dose-response model).

- Step 3: Dose-finding Step – Dose-Response Model Fitting

  Once the significance of a dose-response signal is established, the dose-response profile and the target dose can be estimated using the best-fitting dose-response model. All significant dose-response models will be re-fitted to the data without any parameter assumptions to generate a set of new estimates of the model parameters from the data. The final dose-response model will be obtained via the model with the smallest Akaike Information Criterion (AIC) (the smaller the AIC value the better the model fit). Estimates for the dose groups will be calculated and will be based on the final dose-response model.
  The dose-response curve estimate with the model-based two-sided 95% confidence interval will be presented graphically. In addition, the plot will include the mean responses from the Cox Regression and the associated 95% confidence intervals for each of the studied dose groups.

The target dose(s) suggested by gMCP-Mod based on time to solid food tolerance serves as an important consideration in dose selection. However, the final dose selection will be determined based on the following:

- Efficacy of doses based on the totality of efficacy data
- Safety endpoints including AEs, and SAEs

8.6.2 Definition of Estimand

Time to solid food tolerance through the Day 30 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), 1.0 mg/kg (0.625 mL/kg), or 0.5 mg/kg (0.3125 mL/kg), every 24 hours (±1 hour) for three consecutive days for a total of three infusions.
2. Population: Patients with Acute Pancreatitis and Accompanying Systemic Inflammatory Response Syndrome
3. Variable: Time to solid food tolerance as the number of hours from the date and time of the SFISD for the patient to the first date and time the patient receives a solid meal that is

tolerated, which is defined as eating ≥50% of a low fat, ≥500-calorie solid meal without an increase in abdominal pain or vomiting within the two hours of mealtime.

4. Population level summary: hazard ratio using stratified Cox Regression

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

The null and alternative hypotheses are:

(1) Null hypothesis H0: Time to solid food tolerance curves are the same between Auxora and placebo

(2) Alternative hypothesis H1: Time to solid food tolerance curves are different between Auxora and placebo

8.6.3 Main Analytical Approach

Time to solid food tolerance through Day 30 visit will be analyzed based on the Generalized Multiple Comparison Modelling (gMCP-Mod) approach described in Session 8.6.1.

Time to solid food tolerance is defined as the number of hours from the date and time of the SFISD for the patient to the first date and time the patient receives a solid meal that is tolerated. If the mealtime was not recorded, the mealtime will be estimated using the middle time of the meal serving interval.

Tolerating a solid meal is defined as eating ≥50% of a low fat, ≥500-calorie solid meal without an increase in abdominal pain or vomiting within the two hours of mealtime. In addition, to meet the definition, all subsequent meals during the hospitalization should not show an increase in abdominal pain or vomiting within the two hours of mealtime. If a procedure that could impact food tolerance is performed after a solid meal is initially tolerated that results in subsequent meals not being tolerated, the time of the initial tolerated meal will be considered in the calculation of the time to solid food tolerance. The procedures that impact food tolerance and/or its assessment are defined as (1) a procedure that requires a period of nil by mouth (2) a procedure and/or treatment that causes gastrointestinal dysfunction e.g., cholecystectomy, necrosectomy. The procedures performed after the initial food tolerance will be reviewed before the database lock to identify the procedures that could impact the time to the tolerated solid food. If the subject was not discharged from the hospital during the study and did not tolerate solid food, the time to solid food tolerance will be censored at the latest date and time of the hospital stay.

If a patient is discharged home without tolerating solid food, the daily record of the modified American Neurogastroenterology and Motility Society Gastrointestinal Cardinal Symptom Index Daily Diary (mGCSI-DD) at or after hospital discharge will be used in the calculation of the time to solid food tolerance. The patients who did not tolerate solid food while hospitalized will be considered as tolerating solid food at or after hospital discharge at 8am on the first of three consecutive days where the following 4 criteria are met for the mGCSI-DD:

- none for vomiting
- none or mild for nausea
- none or mild for not being able to finish a normal sized meal for a healthy person
- none or mild for abdominal pain

If the mGCSI-DD is not available at or after discharge in a patient discharged home without tolerating a solid meal, the time to solid food will be censored at 168 hours.

8.6.4 Sensitivity Analysis

The sensitivity analysis is to evaluate the assumptions of using the mGCSI-DD in the primary analysis of patients discharged from the hospital without solid food tolerance. Time to solid food tolerance for patients who could not tolerate solid food during the hospital will be censored to 168 hours if the patient is discharged from the hospital during the study and not tolerating solid food. If the subject was not discharged from the hospital during the study and did not tolerate solid food, the time to solid food tolerance will be censored at the latest date and time of the hospital stay.

8.6.5 Supplementary Analyses

Time to solid food tolerance through the Day 30 visit will be displayed using a Kaplan-Meier estimate and will be compared between the 3 treatment groups and the placebo group using a stratified log-rank test stratified by eCRF recorded sex (male or female) and the risk for organ failure (higher or lower) based on eCRF data.

Supplementary analysis 1: The definition of time to solid food tolerance and the censoring method follow the primary analysis method.

Supplementary analysis 2: The definition of time to solid food tolerance and the censoring method follow the sensitivity analysis method.

8.7 Secondary Endpoint Analyses

8.7.1 Percentage of Patients who Tolerated Solid Food at 48, 72, and 96 Hours from the SFISD and at Discharge

Solid food tolerance will be calculated using the main analytical approach as described in Section 8.6.3.

If the time to first solid food tolerance $\leq$ 48 hours then solid food tolerance at 48, 72 and 96 hours are all yes.

If the time to first solid food tolerance >48 and $\leq$72 hours, then solid food tolerance at 72 and 96 hours are all yes and solid food tolerance at 48 hours is no.

If the time to first solid food tolerance >72 and $\leq$96 hours, then solid food tolerance at 96 hours is yes and solid food tolerance at 48 and 72 hours are all no.

If the time to first solid food tolerance > 96 hours or the patient never tolerates the solid food during the study then solid food tolerance at 48, 72, and 96 hours are all no.

If the time to first solid food tolerance ≤time to hospital discharge, solid food tolerance at discharge is yes, otherwise, solid food tolerance at discharge is no. If a patient was not discharged from the hospital before end of study, solid food tolerance at the end of study is no.

If it is unknown if solid food has been tolerated at a specific time point (time to first solid food tolerance is censored before the specific time point), the solid food tolerance will be imputed as no.

The proportion and 95% CI of patients who tolerated solid food at 48, 72, and 96 hours from the SFISD, at discharge, and at the end of the study will be displayed using a Clopper-Pearson interval and will be compared between each Auxora treatment group and placebo group using a Cochran-Mantel-Haenszel test stratified by the eCRF recorded sex (male or female) and the risk for organ failure (higher or lower) based on eCRF data. Prohibited immunosuppressive medications or immunotherapies and treatment discontinuation will be ignored (treatment policy strategy).

8.7.2 Time to Medically Indicated Discharge

The time to medically indicated discharge is the number of days from the SFISD to the first date of meeting the criteria defining medically indicated discharge:

- the patient tolerates solid food, as defined by the main analytical approach; and
- abdominal pain is controlled or resolved, which will be defined by a reduction in pain, as assessed by the PNRS scale, of ≥50% from the peak level in the first 24 hours and no use of opioids; and
- there is no clinical evidence of an infection requiring continued hospitalization.

In patients who do not meet the criteria for medically indicated discharge when discharged from the hospital, the time to medically indicated discharge will be defined as 8am on the first of three consecutive days where the 4 criteria are met for the mGCSI-DD as defined in 8.6.3. If a patient is discharged not meeting the criteria for medically indicated discharged and does not meet criteria for the mGCSI-DD or has not completed the mGCSI-DD, the time to medically indicated discharge will be censored at 168 hours. If the patient dies while hospitalized, the time to medically indicated discharge will be censored at Day 30.

The time to medically indicated discharge will be summarized using Kaplan-Meier analyses. The medians and quartiles of time-to-events will be summarized for each study group. The time to medically indicated discharge will be compared between each Auxora treatment group and placebo group using stratified log-rank test stratified by sex (male or female) and the risk for organ failure (higher or lower). Prohibited immunosuppressive medications or immunotherapies and treatment discontinuation will be ignored (treatment policy strategy).

The proportion and 95% CI of patients who meet the medical discharge criteria by the 96-hour visit and the date of discharge will be displayed using a Clopper-Pearson interval and will be compared between each Auxora treatment group and placebo group using a Cochran-Mantel-Haenszel test stratified by the eCRF recorded sex (male or female) and the risk for organ failure (higher or lower) based on eCRF data. Prohibited immunosuppressive medications or immunotherapies and treatment discontinuation will be ignored (treatment policy strategy).

The correlation between meeting the medically indicated discharge criteria on the day of discharge with each of the following events, hospital readmission for acute pancreatitis, the development of pancreatic necrosis/necrotizing pancreatitis, and worsening CTSI score at Day 30 from Screening, will be summarized by treatment groups by Matthews correlation coefficient (MCC).

8.7.3 Length of Stay in the Hospital

The number of days patients are in the hospital from the SFISD, including ICU stay and readmissions for any etiology, during the 30 Days of the study will be summarized by treatment groups and compared between each Auxora treatment group and placebo group using an analysis of variance model, which includes treatment group, the stratification factor of the eCRF recorded sex (male or female), and then the stratification factor of the risk for organ failure (higher or lower) based on eCRF data as the fixed effects in the model. The proportion of patients in each quartile of hospital stay (1-7 days, 8-14 days, 15-21 days, and 22-30 days) and the patients who remain hospitalized at Day 30 will also be summarized by treatment groups.

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 30 if the patient dies due to acute pancreatitis (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).

8.7.4 Re-hospitalization for acute pancreatitis (AP) Through Day 30

The date of each re-hospitalization and the reasons for re-hospitalizations will be collected for each patient. The proportion and 95% CI of patients who are re-hospitalized for either acute pancreatitis or the development of pancreatic necrosis/necrotizing pancreatitis through the Day 30 visit will be displayed using a Clopper-Pearson interval and will be compared between each Auxora treatment group and placebo group using a Cochran-Mantel-Haenszel test stratified by the sex (male or female) and the risk for organ failure (higher or lower) based on eCRF data. Prohibited immunosuppressive medications or immunotherapies and treatment discontinuation will be ignored (treatment policy strategy). If the re-hospitalization for acute pancreatitis status is unknown, the re-hospitalization for acute pancreatitis will be imputed as No.

The proportion and 95% CI of patients who are re-hospitalized for other causes through Day 30 will also be analyzed in a similar manner.

8.7.5 Change in Severity of AP by CTSI Score from Screening to Day 30

Independent reviewers will perform a blinded central review of Contrast Enhanced Computed Tomography (CECT) imaging data, or MRI imaging data, obtained at the Screening and Day 30 visits to assess the computed tomography severity index (CTSI). Independent review will be performed by two independent radiologists (primary review) using a consensus methodology.

If the patient undergoes additional CECTs or MRIs at the discretion of the PI or treating physician from randomization to the Day 30 visit, the CTSI data will be listed but not used in the analysis unless the Day 30 visit data are missing. If the Day 30 visit data is missing, the last available unscheduled post-treatment data will be used for the change from baseline at Day 30 analysis based on the LOCF method.

Shift tables will display changes in severity by CTSI score (mild, moderate, severe) from baseline compared to Day 30 by treatment groups.

Proportion of patients with moderate or severe AP by CTSI score at baseline who became mild AP at Day 30, and the proportion of patients with mild AP at baseline who become moderate or severe AP will be displayed and analyzed in the same way as the re-hospitalization for acute pancreatitis through Day 30.

8.7.6 Development of pancreatic necrosis ≥30% and >50%

The development of pancreatic necrosis ≥30% and >50% through Day 30 will be assessed by the independent reviewers in the same way as the CTSI score.

The proportion and 95% CI of patients who developed pancreatic necrosis ≥30% and >50% during the study will be displayed and analyzed in the same way as the re-hospitalization for acute pancreatitis through Day 30. If the Day 30 data is missing, the last available unscheduled post-treatment data will be used as pancreatic necrosis status at Day 30 based on the LOCF method.

The proportion and 95% CI of patients who developed necrotizing pancreatitis, defined as either pancreatic or peripancreatic necrosis, during the study will be displayed and analyzed in the same way as the re-hospitalization for acute pancreatitis through Day 30. If the Day 30 data is missing, the last available unscheduled post-treatment data will be used as pancreatic necrosis status at Day 30 based on the LOCF method.

8.7.7 Persistence of SIRS ≥48 Hours After the SFISD

The persistence of SIRS ≥48 Hours after the SFISD is defined by the presence of at least two of the following four criteria at the first visit 48 hours after the SFISD:

- Temperature < 36°C or > 38°C;
- Heart rate > 90 beats/minute;
- Respiratory rate >20 breaths/minute or arterial carbon dioxide tension (PaCO2) <32 mmHg;

- White blood cell count (WBC) >12,000 mm3, or <4,000 mm3;

The proportion and 95% CI of patients who developed persistence of SIRS ≥48 Hours After the SFISD will be displayed and analyzed in the same way as the re-hospitalization for acute pancreatitis through Day 30. If the patient is discharged before 48 Hours after the SFISD, the last available post-treatment data will be used to define the SIRS status at 48 hours after the SFISD based on the LOCF method.

8.7.8 Incidence, Severity, and Duration of Organ Failure

The incidence, severity and duration of organ failure will be defined by the development of severe respiratory failure or severe renal failure or severe cardiac failure. Severe respiratory failure will be defined as those patients receiving invasive mechanical ventilation (IMV) or those receiving for ≥ 48 hours use of either high flow nasal cannula (HFNC) or non-invasive mechanical ventilation (NIMV). Use of IMV, HFNC, or NIMV for ventilatory support during a procedure, or use of NIMV for the treatment of obstructive sleep apnea, will not be considered as meeting the definition of severe respiratory failure. Severe renal failure will be defined as the initiation of renal replacement therapy. Severe cardiovascular failure will be defined as the use of vasopressor or inotropic support for ≥48 hours. Use of vasopressor or inotropic support during a procedure will not be considered as meeting the definition of severe cardiovascular failure.

The proportion and 95% CI of patients who developed severe respiratory failure or severe renal failure or severe cardiovascular failure will be displayed and analyzed in the same way as the re-hospitalization for acute pancreatitis through Day 30. The proportion of and 95% CI of patients who developed more than one organ failure (multi-organ failure) will be also displayed and analyzed. The mean duration of severe respiratory failure, severe renal failure, and severe cardiac failure will be summarized by treatment groups.

Given the previous results from study CM4620-204, a sensitivity analysis that will combine the subjects from the 1.25 mL/kg and 0.625 mL/kg groups into one combined group and will combine the subjects from the 0.3125 mL/kg and placebo groups into another combined group. The proportion and 95% CI of patients who developed severe respiratory failure, as well as those who developed multiorgan failure, in the two combined groups will be displayed and analyzed in the same way as the re-hospitalization for acute pancreatitis through Day 30.

A gMCP-Mod analysis of severe respiratory failure will be conducted in the same three steps method as Session 8.6.1.

- Step 1: Conventional Step – Odds Ratio Between Each Dose Versus Placebo

The rate of severe respiratory failure will be analyzed using the logistic regression with treatment, sex (male or female) and the risk for organ failure (higher or lower) as covariates to obtain treatment effects and corresponding variance covariance matrix.

- Step 2: Candidate Models and Multiple Testing Procedure

The optimal contrast based on Candidate models for Auxora dose for the logit of rate of severe respiratory failure follows the following table

Model	Contrast coefficients for dose			
	Placebo	0.5 mg/kg	1.0 mg/kg	2.0 mg/kg
Emax	0.866	-0.264	-0.293	-0.308
sigEmax	0.663	0.300	-0.424	-0.539

Note: Candidate model of Emax model is based on the assumption of parameter of ED50=0.026 mg/kg. Candidate model of sigEmax mode is based on the assumption of parameters of ED50=0.606 mg/kg and h=4.3

- Step 3: Dose-finding step – Dose-Response Model Fitting

The procedure is the same as the step 3 of Session 8.6.1.

The proportion and 95% CI of patients who developed new onset persistent respiratory failure or new onset persistent renal failure or new onset persistent cardiovascular failure will be displayed and analyzed in the same way as rehospitalization for acute pancreatitis through Day 30. New onset is defined as organ failure that was not present at randomization. Persistent is defined as ≥48 hours. Respiratory failure, renal failure, and cardiovascular failure are defined by a score of ≥2 on the modified Marshall scoring system. Patients with a known medical history of chronic kidney disease will be excluded from the analysis of new onset renal failure.

8.7.9 Mortality by Day 30

Mortality is recorded at the Day 30 follow-up assessment. The proportion and 95% CI of patients who died by Day 30 will be displayed and analysed in the same way as the re-hospitalization for acute pancreatitis through Day 30. If the Day 30 vital status is unknown and the patient has been discharged, the patient will be treated as alive.

8.7.10 Change in Pain Score

Pain numeric rating scales are recorded through the Day 10 visit and at the Day 30 visit. The mean score at each visit and mean changes from baseline to each post-baseline visit of the pain numeric rating scale will be summarized by the treatment group.

8.8 Exploratory Endpoints

The development of infected pancreatic necrosis, and of sepsis will be summarized by the number of patients and treatment group.

Mean changes in total mGCSI-DD score from screening to discharge and to Day 30, IL-6 levels, and urine neutrophil gelatinase-associated lipocalin (NGAL) levels to each post-baseline visit will be summarized by the treatment groups.

The mean change in IL-6 levels in patients with a baseline level ≥150 pg/mL to the 72-hour visit will be summarized by the treatment group. The change in urine NGAL in patients with a baseline level ≥150 ng/mL to the 72-hour visit will be summarized by the treatment groups.

Shift tables will display the number of patients with baseline IL-6 levels >1000 pg/mL, 150-1000 pg/mL at 72 hours by the treatment groups.

8.9 Proof-of-Concept Composite Endpoint Analysis

Finkelstein-Schoenfeld Analysis

The proof-of-concept efficacy analysis uses a hierarchical combination applying the method of Finkelstein-Schoenfeld (Finkelstein 1999) to the following:

1. All-cause mortality at Day 30 Visit, and
2. Severe respiratory failure through the Day 30 Visit, and
3. Pancreatic necrosis ≥30% through the Day 30 Visit, and
4. Rehospitalization for AP or pancreatic necrosis/necrotizing pancreatitis through the Day 30 Visit, and
5. Time to solid food tolerance within 30 Days of the SFISD

The test score will be computed using a pairwise ranking procedure. The test statistic will be based on the sum of these scores and was stratified by the eCRF recorded sex (male or female) and then the stratification factor of the risk for organ failure (higher or lower) based on eCRF data.

Within each stratum, each subject will be compared to every other subject in a pair-wise manner. The pair-wise comparison assigned a +1 to the "better" subject and a -1 to the "worse" subject. All-cause mortality will be given higher priority in the calculation and only when two given subjects are alive at Day 30 when evaluating on all-cause mortality, then persistent severe respiratory failure, pancreatic necrosis >30%, rehospitalization for AP, and time to solid food tolerance would be used in the comparison in sequence.

- If one subject is alive and the other is not, the live subject receives a +1 and the deceased one a -1.
- If both subjects died, then the comparison is a tie.
- If both subjects are alive, then the comparison evaluates persistent severe respiratory failure to assign scores. The subject with persistent severe respiratory failure receives a -1 while the other without persistent severe respiratory failure receives +1.
- If both patients have, or do not have, persistent severe respiratory failure, then the comparison evaluates pancreatic necrosis ≥30%. The subject with pancreatic necrosis ≥30% receives a -1 while the other without pancreatic necrosis 30% receives +1.

- If both subjects have, or do not have, pancreatic necrosis ≥30%, then the comparison evaluates rehospitalization for acute pancreatitis. The subject with rehospitalization for acute pancreatitis receives a -1 while the other without rehospitalization for acute pancreatitis receives +1.
- If both subjects have, or do not have, rehospitalization for acute pancreatitis, the comparison uses time to solid food tolerance to assign scores. The subject with the longer time to solid food tolerance receives a -1 while the other with shorter time to solid food tolerance receives +1.

No imputation will be done for missing cases for the proof-of-concept composite efficacy analysis based on the Finkelstein-Schoenfeld method. In case when one subject is censored and all-cause mortality, persistent severe respiratory failure, pancreatic necrosis ≥30%, rehospitalization for AP, or time to solid food tolerance are unknown, the unknown status will be used in calculating the score and the score will be considered 0.

The proposed test is a score test based on the sum of the scores for the treated group. The test statistic can be written as

$$T = \sum_k \sum_{i \in A_k} D_i U_i$$

A_k is the set of the subjects in the kth strata and U_i is the score calculated within each stratum. D_i=1 for subjects in Auxora group and D_i=0 for subjects in placebo group. K=1,2,3,4 for four strata. The variance is written as

$$V = \sum_k \frac{m_k(n_k - m_k)}{n_k(n_k - 1)} \left(\sum_{i \in A_k} U_i^2\right)$$

M_k is the total number of Auxora subjects in the kth stratum. N_k is the total number of subjects in the kth stratum.

The hypothesis will be tested by comparing T/ $\sqrt{V}$.

Win Ratio and Confidence Interval

The win ratio method involves a comparison of each patient in an active treatment group with each patient in the placebo group to determine whether the active treatment patient had a better outcome (a "win"), a worse outcome (a "loss"), or an identical outcome (a "tie"). Each comparison begins with the death outcome and proceeds through the hierarchically ordered set of outcomes described below until the first outcome is identified for which the outcomes differ between the patients; this outcome is used to determine the "win" or "loss." The win ratio is the total number of wins divided by the total number of losses, and a win ratio >1 indicates better outcomes with active treatment.

Categories (a), (c), (e), (g), and (i) represent Auxora wins based on all-cause mortality, persistent severe respiratory failure, pancreatic necrosis >30%, rehospitalization for AP, and time to solid

food tolerance sequentially. Similarly, categories (b), (d), (f), (h), and (j) represent placebo wins. Category (k) represents ties, pairs where subjects were not able to be differentiated.

(a) Death on placebo, but alive on Auxora at Day 30
(b) Death on Auxora, but alive on placebo at Day 30
(k) Death on placebo, death on Auxora at Day 30. No further comparisons will be performed.

If alive on placebo and alive on Auxora at Day 30:
(c) Severe respiratory failure on placebo but not on Auxora
(d) Persistent severe respiratory failure on Auxora but not on placebo

If a tie for severe respiratory failure:
(e) Pancreatic necrosis ≥30% on placebo but not on Auxora
(f) Pancreatic necrosis ≥30% on Auxora but not on placebo

If a tie for pancreatic necrosis ≥30%:
(g) Rehospitalization for AP on placebo but not on Auxora
(h) Rehospitalization for AP on Auxora but not on placebo

If a tie for rehospitalization for AP:
(i) Earlier time to solid food tolerance on Auxora
(j) Earlier time to solid food tolerance on placebo

If a tie for rehospitalization for AP
(k) Tie for all outcomes

The overall win ratio will be calculated by adding (a) + (c) + (e) + (g) + (i) for all 4 strata and dividing it by the sum of (b) + (d) + (f) + (h) + (j) across all 4 strata.

The standard error of log (Win Ratio) can be calculated as:
SE (Log Win Ratio) = log (Win Ratio) / Z score from FS method
The confidence interval can be computed as follows:
95% CI for Win Ratio = exp (log (Win Ratio) +/- 1.96 * SE (Log Win Ratio))

The Win Ratio with 95% CI and p-value based on Finkelstein-Schoenfeld method will be used to compare each Auxora group versus placebo group.

The proof-of-concept analysis will perform a sensitivity analysis that will combine the subjects from the 1.25 mL/kg and 0.625 mL/kg groups into one combined group and combine the subjects from the 0.3125 mL/kg and placebo groups into a second combined group. The two combined groups will be compared using the Finkelstein-Schoenfeld method and Win Ratio method (Finkelstein 1999). This grouping method assumes that the 1.25 mL/kg and 0.625 mL/kg groups have the similar treatment effect on persistent severe respiratory failure. If this assumption is not valid after the primary analysis of the time to solid food tolerance after the database lock, the grouping method may be adjusted.

8.10 Safety Endpoint Analyses

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

8.10.1 Adverse Events

Adverse events (AEs) 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 8.1.5 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 Related', 'Possibly Related', 'Probable Related', 'Definite Related'. AEs classified as 'Possibly Related', 'Probable Related', or 'Definite Related'' 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 eCRF.

8.10.1.1 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 Total 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 SOC 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 TEAEs and TEAEs leading to discontinuation of study treatment will be generated as well.

8.10.1.2 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

8.10.1.3 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

8.10.2 Vital Signs

Vital signs will include temperature, systolic and diastolic blood pressure, respiratory rate (breaths per minute), heart rate (beats per minute) and QTc (milliseconds). 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 >130 bpm
- Pulse rate <45 bpm
- Diastolic blood pressure >100 mmHg
- Systolic blood pressure >155 mmHg
- Systolic blood pressure <80 mmHg
- Respiration rate >25 breaths/min
- Body temperature ≥39 °C
- QTc >500 milliseconds

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.

8.10.3 Laboratory Analyses

Local laboratory test results will be used in the summary. CBC with differential, platelets, serum chemistries, and serum procalcitonin will be summarized.

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:

- Shift from baseline category to the worst category
- Shift from baseline category to the last category

8.10.4 Arterial Blood Gas

If an arterial blood gas (ABG) has been performed as supportive care, the time of the draw and following results will be listed: the pH, PaO_2, $PaCO_2$, SaO_2 and FiO_2.

8.10.5 SpO_2/FiO_2

SpO_2 and FiO_2 parameters, and imputed P/F ratios will be listed.

8.11 PK Assessments

PK analysis will be performed as a separate analysis.

8.12 Subgroup Analyses

Subgroup analyses will be performed to explore how time to solid food tolerance through Day 30 visit and proof-of-concept endpoint 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, Finkelstein-Schoenfeld Method, and Win Ratio will be performed by subgroup levels of the baseline variables listed below:

- Age (<50 vs ≥50)
- Race (White vs Asian vs Other)
- Sex (Male vs Female)
- Etiology (Alcohol vs Other)
- SIRS (Two criteria vs > Two criteria)
- Risk For Organ Failure (High vs Low as determined by values entered in eCRF)
- Hypoxemia (imputed P/F ≤360 vs >360)
- HCT (High: HCT ≥44% for men or ≥40% for women vs Low: HCT <44% for men or <40% for women)
- Balthazar Score on Screening CECT by central reading (A or B or C, vs D or E)
- WBC (>12,000 mm^3 or < 4,000 mm^3 vs ≤12,000 mm^3 and ≥4,000 mm^3)
- Patients enrolled in India vs. in the United States of America

To increase the number of subjects in each comparison group of the subgroup analysis, the subgroup analysis will combine the subjects from the 1.25 mL/kg and 0.625 mL/kg groups into one pooled Auxora group. This pooled Auxora group will be compared with the placebo group. This grouping method assumes that the 1.25 mL/kg and 0.625 mL/kg groups have the similar treatment effect. If this assumption is not valid after the primary analysis of the time to solid food tolerance, the grouping method will be adjusted.

9. INTERIM ANALYSIS

9.1 Interim Safety Analyses

Scheduled periodic review meetings will be conducted by an independent data monitoring committee (IDMC) overseeing the development program of Auxora. An IDMC charter will govern the IDMC and will describe the scope of responsibilities of the IDMC. If the IDMC

recommends alteration of the dosing regimen because of safety issues, the FDA and other regulatory agencies will be notified as appropriate. The IDMC responsibilities include protecting the safety of the study patients and making recommendations to CalciMedica concerning the conduct of the study.

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 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. SUMMARY OF CHANGES TO STATISTICAL ANALYSIS PLAN VERSION 1.0

Table 3. Summary of changes from statistical analysis plan version 1.0

#	**Nature of change from Statistical Analysis Plan 1.0**	**Rationale**
1	N/A	

13. SUMMARY OF CHANGES TO PROTOCOL PLANNED ANALYSIS

Table 4. Summary of changes from protocol planned analysis

#	**Nature of change from Protocol Planned Analysis**	**Rationale**
1	The analysis of the secondary endpoints of length of stay in the ICU for patients admitted to the ICU.	Length of stay in the ICU will be integrated into length of hospital stay.

14. BIBLIOGRAPHY

1. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics 2005; 61:738–748.

2. Pinheiro J, Bornkamp B, Glimm E, Bretz F. Model-based dose finding under model uncertainty using general parametric models. Statistics in Medicine 2014; 33:1646-1661.

3. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. Stat Med. 1999 Jun 15;18(11):1341-54. doi: 10.1002/(sici)1097-0258(19990615)18:11<1341::aid-sim129>3.0.co;2-7. PMID: 10399200.