CONFIDENTIAL

IDN-6556 (Emricasan)

IDN-6556-14 STATISTICAL ANALYSIS PLAN (VERSION 1)

Protocol Title A Multicenter, Randomized, Double-Blind, Placebo-Controlled Trial of Emricasan, an Oral Caspase Inhibitor, in Subjects with Non-Alcoholic Steatohepatitis (NASH) Cirrhosis and Severe Portal Hypertension

Sponsor Conatus Pharmaceuticals Inc.
16745 W. Bernardo Drive, Suite 200
San Diego, CA 92127

Statistician James M Robinson, M.S.
Senior Director, Biostatistics & Data Management

Version 2.0 10 JUL 2018

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

AE	Adverse Event
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical Classification
AUDIT	Alcohol Use Disorders Identification Test
BID	bis in die, twice daily
BMI	Body Mass Index
CFB	Change from Baseline
CI	Confidence Interval
CLDQ	Chronic Liver Disease Questionnaire
C-P	Child-Pugh
$cPDR_{30}$	Cumulative Percentage Dose Recovery of ^{13}C-methacetin 30 min after ingestion of test substrate
DMC	Data Monitoring Committee
eCRF	Electronic Case Report Form
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
E_0	Basal study drug effect
ED_{50}	Median effective dose
ELF	Enhanced Liver Fibrosis
E_{max}	Maximum study drug effect
FAS	Full Analysis Set
FDA	Food and Drug Administration
HVPG	Hepatic Venous Pressure Gradient
ICF	Informed Consent Form
IWRS	Interactive Web Randomization System
LSMean	Least-Square Mean
MBT	^{13}C-methacetin breath test
MCS	SF-36 mental-component summary score
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-stage Liver Disease
MI	Multiple Imputation
Na	Sodium
NASH	Non-alcoholic steatohepatitis
NSBB	Non-selective beta-blockers
PCS	SF-26 physical-component summary score
PD	Pharmacodynamics
PK	Pharmacokinetics
PPS	Per-protocol set
QTcB	Bazett's QT correction

QTcF	Fridericia's QT correction
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SF-36	Short form 36 questionnaire
SOC	System Organ Class
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment-Emergent Adverse Event
ULN	Upper Limit of Normal range
WHODRL	World Health Organization Drug Reference List

1 INTRODUCTION

IDN-6556-14 Protocol describes the collection and analysis of clinical study data to evaluate the use of IDN-6556 (emricasan), an oral caspase inhibitor, in subjects with non-alcoholic steatohepatitis (NASH) cirrhosis and severe portal hypertension. This document provides details of the statistical analyses to be performed. All decisions regarding handling of data for reporting results will be determined prior to database lock and will be described in the relevant sections of this statistical analysis plan (SAP).

2 STUDY OBJECTIVES

2.1 PRIMARY OBJECTIVE

The primary objective of this study is to assess whether emricasan compared to placebo leads to a mean decrease in hepatic venous pressure gradient (HVPG) at Week 24 in subjects with NASH cirrhosis and severe portal hypertension.

2.2 SECONDARY OBJECTIVES

Secondary objectives of this study are:

- To assess the safety and tolerability of emricasan
- To evaluate the dose response of emricasan on portal pressure as assessed by HVPG at Week 24
- To assess whether emricasan compared to placebo improves HVPG response at Week 24 using a 20% reduction from baseline response definition
- To assess whether emricasan compared to placebo decreases mechanism specific (caspase 3/7) and non-specific (alanine aminotransferase [ALT]) biomarkers at Weeks 24 and 48

2.3 EXPLORATORY OBJECTIVES

Exploratory objectives of this study, as defined by the study protocol, are:

- To assess whether emricasan compared to placebo improves HVPG response at Week 24 using response criteria of 10% reduction from baseline response definition
- To assess whether emricasan compared to placebo improves liver function and prognosis at Weeks 24 and 48 as assessed by model for end-stage liver disease (MELD) and Child-Pugh (C-P) scores (change in score, progression and regression)
- To assess whether emricasan compared to placebo improves biochemical and functional biomarkers (cCK18/M30, flCK18/M65, aspartate aminotransferase (AST), total bilirubin, international normalized ratio (INR), and albumin) at Weeks 24 and 48
- To assess whether emricasan compared to placebo improves fibrosis markers at Weeks 24 and 48
- To assess whether emricasan compared to placebo improves health-related quality of life at Weeks 24 and 48
- To assess whether emricasan compared to placebo decreases development of decompensation or worsening of decompensation at Weeks 24 and 48
- To assess whether emricasan compared to placebo improves liver metabolic function at Weeks 24 and 48 as assessed by methacetin breath test (at select sites)
- To assess whether emricasan compared to placebo improves liver stiffness at Weeks 24 and 48 as assessed by transient elastography (FibroScan®) (at select sites)

3 STUDY DETAILS

3.1 STUDY DESIGN

This is a multicenter, double-blind, randomized, placebo-controlled, dose-response study to evaluate the safety and efficacy of emricasan in improving portal hypertension in subjects with NASH cirrhosis and severe portal hypertension (defined as HVPG ≥12 mmHg). Subjects can have compensated (at least 60% of subjects but no more than 75%) or decompensated cirrhosis with no more than 1 prior significant decompensating event and must be currently clinically stable on stable standard therapy (see Inclusion/Exclusion criteria). Randomization will be stratified by compensated vs. decompensated status at baseline as well as use of non-selective beta-blockers (NSBB) or not. Subjects who are otherwise eligible will undergo the HVPG procedure as the last qualifying procedure prior to Day 1.

The study treatment duration will be up to 48 weeks, including an initial 24-week randomized treatment phase with follow-up HVPG at Week 24 (primary endpoint) and an additional 24-week treatment phase (continuing the same study drug treatment as initially randomized). Subjects completing the initial 24-week randomized treatment phase will be re-consented for the additional 24-week treatment phase. HVPG measurements will be performed at screening and Week 24, with additional placebo controlled safety follow-up and exploratory efficacy assessments through Week 48. Subjects will complete a final follow-up visit approximately 2 weeks after the end of treatment (i.e., at Week 26 if completing the initial 24-week randomized treatment only, or at Week 50 if completing the 48week treatment).

Subjects with baseline HVPG ≥12 mmHg may experience progression of cirrhosis during the 48-week study. Given that the primary endpoint for this study is HVPG rather than a clinical event (or clinical composite), subjects who have a decompensating event or worsened decompensation during the study will not be required to withdraw and can remain in the study unless the investigator feels the subject is not stable or the subject requires a transjugular intrahepatic portal shunt or other portosystemic bypass procedure, which would confound the HVPG assessment. Subjects who progress to Child-Pugh class C will be discontinued from study drug but are expected to remain in the study and complete all planned study visits. All subjects who wish to discontinue study drug treatment should remain in the study and complete all planned study visits, in order to minimize missing data and protect the integrity of the study results. Subjects who withdraw consent for participating in the study should discontinue study visits. Subjects for whom the investigator deems that it is not in the best interest of the subject to continue participating in the study may also discontinue study visits.

Subjects will be randomized in a 1:1:1:1 ratio to emricasan 50 mg BID, emricasan 25 mg BID, emricasan 5 mg BID or matching placebo BID. For each subject, the study will consist of:

- Screening period of up to 6 weeks
- Double-blind treatment period of 24 weeks, single-blind after Week 24 and up to Week 48 (i.e., Investigative site staff, subjects, and third-party vendors will remain blinded)
- A follow-up visit 2 weeks after the Week 24 or Week 48 (or early termination) visit

The duration of each subject's participation is intended to be approximately 32 or 56 weeks.

The schedule of events is provided in IDN-6556-14 Protocol, Appendix I.

There will be two analyses conducted for this study, a primary analysis and a final analysis. The primary analysis to be conducted after all randomized subjects have either discontinued prior to week 24 or completed the Week 24 planned study visit. This primary analysis will be considered as a database freeze and will include all data collected at the time of the database freeze (i.e., will include data collected after the Week 24 study visit). The final analysis will include all subjects and all completed study visits. The clinical study report will be written after the final analysis is conducted.

3.2 STUDY POPULATION

The study population is subjects with NASH cirrhosis and severe portal hypertension (defined as HVPG ≥12 mmHg). Subjects can have compensated (at least 60% of subjects but no more than 75%) or decompensated cirrhosis with no more than 1 prior significant decompensating event and must be currently clinically stable on stable standard therapy (see Inclusion/Exclusion criteria).

3.3 RANDOMIZATION

Subjects will be randomized in a 1:1:1:1 ratio to emricasan 50 mg BID, emricasan 25 mg BID, emricasan 5 mg BID or matching placebo BID. The assignment to emricasan or placebo will be performed randomly. The randomization schedule will be generated using a validated randomization program and verified for accuracy using strict quality control procedures.

The assignment of randomization number and treatment assignment will be centrally coordinated through the study's Interactive Web Randomization System (IWRS). Randomization numbers will be assigned by the IWRS.

The randomization will be stratified by 2 factors:

- Compensated vs. decompensated status at baseline
- Use of non-selective beta-blockers (NSBB)

3.4 BLINDING

This clinical trial is a double-blind study. The sponsor will be unblinded after the completion of the Week 24 study visit with sites and subjects remaining blinded until the completion of the final study visit. Information on individual treatment assignments will not be allowed outside of the sponsor until the final analysis conducted after all subjects complete their final visit (i.e., when the last active subject has either discontinued the study or completed their Week 50 study visit).

Investigators will be able to unblind subjects through the IWRS when it is medically imperative to know whether a subject is receiving emricasan or placebo, such as in the event of an adverse event (AE) that the Investigator feels cannot be adequately treated without knowing the identity of the study drug. Every effort must be made to contact the Medical Monitor to discuss the case before breaking the blind, or if in an emergency, as soon as possible thereafter (no later than 24 hours after emergency unblinding) to inform the Medical Monitor that unblinding was performed but without disclosing the actual treatment assignment. Investigators should make arrangements to ensure that access to the secure internet site (i.e., individual user name and password) is

maintained in strict confidence to prevent a compromise of subject blinding by non-study or unauthorized individuals.

The Sponsor may access the randomization codes for subjects with potential suspected unexpected serious adverse reactions (SUSARs) for the purpose of regulatory reporting or for the purpose of evaluating an emergent safety issue. In such an event, the Sponsor will document the rationale, circumstances, and the person or persons being informed about the unblinding.

If the blind for a subject is broken (by the Investigator or the Sponsor), an entry must be made in the electronic data capture system that contains the reason that the blind was broken and the name of the person contacted at the Sponsor or designee.

3.5 ANALYSIS WINDOWS

Although study visits are designated on the electronic case report forms (eCRFs), data from all visits (i.e., scheduled and unscheduled) will be classified into analysis visit windows for purposes of analysis. Target dates for each subject's scheduled visit throughout the study will be derived based on their first dose date (i.e., Day 1). Each study visit date will then be classified into the analysis visit windows as listed in Table 1. The priority within an analysis window is that the regular scheduled visit will be used. In the event any data assessments are missing from the regular scheduled visit, data from the unscheduled visit closest to the target date will be used for each missing assessment. Baseline values are defined in Section 7.

In addition, post-baseline HVPG values will also be assigned to a Week 24 flag based on the HVPG date and will be classified as being Week 24 if the HVPG date is collected within ±8 weeks (i.e., 56 days, inclusive) from the target date for the Week 24 visit

Table 1. Analysis Visit Windows

Visit	Target Study Day	Analysis Window
Screening	<Day -1	≤ Day -1
Day 1	Day 1	Day 1
Week 4	Day 28	> Day 1 and ≤ Day 42
Week 8	Day 56	> Day 42 and ≤ Day 70
Week 12	Day 84	> Day 70 and ≤ Day 98
Week 16	Day 112	> Day 98 and ≤ Day 126
Week 20	Day 140	> Day 126 and ≤ Day 154
Week 24	Day 168	> Day 154 and ≤ Day 196
Week 26[a]	Day 182	> Day 182
Week 32	Day 224	> Day 196 and ≤ Day 252
Week 40	Day 280	> Day 252 and ≤ Day 308
Week 48	Day 336	> Day 308 and ≤ Day 350
Week 50[b]	Day 350	> Day 350

[a]Week 26 analysis window only applies to subjects that do not consent to the Weeks 28 – 48 study visits.

[b]Week 50 analysis window only applies to subjects that consented to the Weeks 28 – 48 study visits.

4 PLANNED SAMPLE SIZE DETERMINATION

It is planned to randomize 240 subjects (60 subjects per group) to receive 1 of 3 emricasan doses or placebo. Subjects will be randomly assigned to receive either emricasan 50 mg BID, emricasan 25 mg BID, emricasan 5 mg BID, or matching placebo in a 1:1:1:1 ratio. Assuming a 20% attrition rate, it is expected that approximately 192 subjects (48 subjects per group) will have a Week 24 HVPG assessment.

This sample size will provide 81% power to detect a statistically significant difference, using a 2-sided test with alpha of 0.05, between at least 1 emricasan treatment group and placebo in the mean change from baseline in HVPG. This calculation assumes a mean difference between an active treatment group and placebo of 3 mmHg and a sample standard deviation of 4.5 mm Hg. This calculation also applied a one-step Dunnett's adjustment for comparing 3 active treatment groups with placebo.

Based on data in the literature and discussion with key experts in the field, a clinically meaningful difference in the change from baseline in HVPG is approximately 2.5-3 mmHg, depending on the baseline value (D'Amico 2006). Subjects in Study IDN-6556-11 (open-label pilot study evaluating the effect of emricasan on HVPG for 28 days) with a baseline HVPG ≥12 mm Hg had a sample standard deviation of 4.05 mm Hg.

The sample size and power were calculated using the IndividualPower SAS macro provided in "Multiple Comparisons and Multiple Tests using SAS" (Westfall 2003).

5 STUDY ENDPOINTS

5.1 PRIMARY EFFICACY ENDPOINT

The primary endpoint for this study is the change from baseline (CFB) at Week 24 in HVPG. The change from baseline will be calculated as HVPG at Week 24 minus HVPG at baseline (HVPGW24-HVPGBL). Hence, a negative value will represent a decrease in HVPG at Week 24.

The HVPG measurement is the difference between wedged (or occluded) hepatic vein pressure and free hepatic vein pressure and represents the gradient between the portal vein and the intra-abdominal inferior vena cava pressure, with normal HVPG being 3-5 mmHg. It is a measure of intrahepatic sinusoidal pressure that correlates directly with direct measurements of portal pressure in cirrhosis.

HVPG will be measured at the Screening and Week 24 visits. Original tracings of the pressure measurements will be kept at the site as part of source documentation for the study, and copies will be forwarded to the Central Reader for evaluation.

5.2 SECONDARY EFFICACY ENDPOINTS

The CFB at each study visit for all continuous secondary endpoints will be calculated as the post-baseline value minus the baseline value. Hence, a negative value will represent a decrease at the post-baseline visit.

5.2.1 HVPG Response 20% Reduction

HVPG response is defined as having at least a 20% reduction at Week 24 (i.e., pCFB HVPG $\leq$ -20%) from baseline in HVPG. This response variable will be used as a secondary endpoint in this study.

5.2.2 Caspase 3/7

Caspase 3/7 data will be log-transformed for purposes of analysis. The CFB on the log-transformed values at each study visit for the mechanistic biomarker caspase 3/7 will be derived and analyzed, with Week 24 and 48 considered as secondary endpoints for this study.

Back transformation of the results will be conducted to provide percent relative change on the original scale.

5.2.3 ALT

Measurements of ALT will be collected at each study visit. ALT data will be log-transformed for purposes of analysis. The CFB on the log-transformed value at each study visit for the functional biomarker ALT will be derived and analyzed, with Week 24 and 48 considered as secondary endpoints for this study.

Back transformation of the results will be conducted to provide percent relative change on the original scale.

5.3 EXPLORATORY ENDPOINTS

The CFB at each study visit for all continuous exploratory endpoints will be calculated as the post-baseline value minus the baseline value. Hence, a negative value will represent a decrease at the post-baseline visit.

5.3.1 HVPG Response 10% Reduction

HVPG response is defined as having a decrease of at least 10% from baseline in HVPG (i.e., pCFB HVPG ≤ -10%), and will be summarized at Week 24.

5.3.2 Model for End-Stage Liver Disease Score, Regression, and Progression

The MELD is a scoring system for assessing the severity of chronic liver disease and short-term mortality. An earlier version of the formula for MELD used the subject's values for serum bilirubin, serum creatinine, and the international normalized ratio (INR) for prothrombin time and was calculated according to the following formula (rounded to the nearest whole number):

Previous MELD =3.78[ln total bilirubin (mg/dL)] + 11.2[ln INR] + 9.57[ln serum creatinine (mg/dL)] + 6.43

where ln is the natural logarithm, mg/dL is milligrams per deciliter

If the patient was dialyzed twice within the last 7 days, then the value for serum creatinine used was set to 4.0. Any value less than one is given a value of 1 (i.e. if bilirubin was 0.8, a value of 1.0 was used) to prevent the occurrence of scores below 0 (the natural logarithm of 1 is 0, and any value below 1 would yield a negative result).

In January 2016, the Organ Procurement and Transplantation Network updated the MELD score to include sodium for the purpose of liver organ allocation. Thus, the MELD score is now calculated based on the previous MELD score (provided above) as well as sodium (Na) according to the following formula (which will be used for this study):

$$\text{MELD} = (\text{previous MELD}) + 1.32 \times (137-\text{Na}) - [0.033 \times (\text{previous MELD}) \times (137-\text{Na})]$$

Sodium values <125 mmol/L will be set to 125, and values >137 mmol/L will be set to 137.

The exploratory endpoint of change from baseline in MELD score at Weeks 24 and 48 will be derived. CFB on the log-transformed value will also be derived and explored.

Regression of MELD will be derived as having a decrease of at least 2 points from baseline in MELD score (i.e., CFB MELD ≤-2) at Weeks 24 and 48. Progression of MELD will be derived as having an increase of at least 4 points from baseline in MELD score (i.e., CFB MELD ≥4) at Weeks 24 and 48.

The components of MELD will be measured at each study visit and the MELD scores will be derived for each study visit by the central laboratory.

Baseline MELD is defined as the last observed value up to and including Day 1.

5.3.3 Child-Pugh Score, Regression, and Progression

The C-P score is used to assess the prognosis of chronic liver disease and is calculated as the sum of the following five component scores: total bilirubin, serum albumin, ascites, hepatic encephalopathy, and INR. Each component is scored with values of 1 to 3 points with 3 indicating the greatest severity. This provides a range for the C-P score of 5 to 15 points and this score is used to determine the C-P classification of A, B, or C, with C being the most severe (Table 2 and Table 3).

Table 2. Child-Pugh Score by Parameter

Parameter	Child-Pugh Score		
	1	**2**	**3**
Total Bilirubin (mg/dL)	<2.0	2.0-3.0	>3.0
Albumin (g/dL)	>3.5	2.8-3.5	<2.8
Ascites	Currently with no ascites or ascites only detectable on imaging (grade 1) and not on diuretics	Currently ascites only detectable on imaging (grade 1) and treated with diuretics OR with overt ascites (grade 2 or 3) that is diuretic responsive	Currently with overt ascites (grade 2 or 3) that is not diuretic responsive
Hepatic Encephalopathy	Currently with no overt HE and not on specific therapy (Rifaxmin)	Currently with no overt HE and no asterixis on treatment with specific therapy OR currently with grade 2 HE (confused but talking and/or with asterixis) not on treatment with specific therapy	Currently with grade 2 HE despite treatment with specific therapy or currently with grade 3 HE or grade 4 HE
INR	<1.7	1.7-2.3	>2.3

Table 3. Child-Pugh Classification

Child-Pugh Classification	Child-Pugh Score
Child-Pugh A	5-6
Child-Pugh B	7-9
Child-Pugh C	10-15

The exploratory endpoint of change from baseline in C-P at Weeks 24 and 48 will be derived. Regression of C-P will be derived as having a decrease of at least 2 points from baseline in C-P score (i.e., CFB C-P $\leq$ -2) at Weeks 24 and 48. Progression of C-P will be derived as having an increase of at least 2 points from baseline in C-P score (i.e., CFB C-P $\geq$ 2) at Weeks 24 and 48.

The components of C-P will be measured at and the C-P score will be derived for each study visit.

5.3.4 Biochemical and Functional Biomarkers

Biochemical and functional biomarker measurements will be collected at each study visit. Exploratory biochemical and functional biomarkers include cCK18/M30, flCK18/M65, AST, total bilirubin, INR, and albumin. These measurements will all be log-transformed for purposes of analysis. The log-transformed CFB at each study visit will be derived for each biomarker.

Back transformation of the results will be conducted to provide percent relative change on the original scale.

5.3.5 Fibrosis Markers

Exploratory fibrosis markers include the Enhanced Liver Fibrosis (ELF) panel (based on hyaluronic acid, P3NP, TIMP-1), ferritin, and LOX-L2. The CFB at Week 24 and Week 48 for each fibrosis marker will be derived. Fibrosis marker measurements will be collected at the Day 1 and Week 24 visits.

5.3.6 CLDQ

The Chronic Liver Disease Questionnaire (CLDQ) was developed as an evaluative instrument to measure longitudinal change in health status within individuals with chronic liver disease. In addition to measuring both physical and mental health, the instrument was designed to be a disease-specific tool for assessing areas of function important to patients with chronic liver disease. The instrument contains 29 items within six domains including abdominal symptoms, fatigue, systemic symptoms, activity, emotional function and worry. A Likert scale response format is used for all items ranging from 1 (most impairment) to 7 (least impairment). Scoring of the questionnaire is performed by dividing each domain score by the number of items per domain. Overall CLDQ score is obtained by adding scores for each item and dividing by the total number of items (n=29).

The change from baseline in the overall CLDQ score and each of the 6 domains will be derived for each visit. The change from baseline at Weeks 24 and 48 will be considered as the exploratory endpoints for CLDQ.

5.3.7 SF-36 Quality of Life

The SF-36 consists of 36 items, 35 of which are aggregated to evaluate eight dimensions of health: physical function, bodily pain, general and mental health, vitality, social function, and physical and emotional role functioning. Scores on the eight subscales are aggregated to derive the physical-component summary score (PCS) and the mental-component summary score (MCS). Scores of each question can range from 0 to 10, and the summary and subscale scores can range from 0 to 100, with higher scores indicating better health.

The change from baseline of each component summary and subscales will be derived for each visit. The change from baseline at Weeks 24 and 48 will be considered as the exploratory endpoints for the component summary and subscales.

5.3.8 Clinical Outcome Events (Decompensation)

Clinical outcome events will be defined based on whether the subject has compensated cirrhosis or decompensated cirrhosis at study entry, and for the latter, depending on what was the prior decompensating event. Subjects may have more than 1 clinical outcome event. Subjects who

have compensated cirrhosis on study entry and have a decompensating event could meet criteria for an additional decompensating event(s) if they remain in the study.

Qualifying clinical outcome events, listed in Table 4 below, will be considered as development of decompensation or worsening of decompensation. Investigators will decide if a subject qualifies for an event as defined in Table 4 and record their assessment of a qualifying clinical outcome event on the eCRF.

Whether or not a subject is classified as having developed decompensation or worsening of decompensation and the number of decompensation events will be exploratory endpoints for this study.

Table 4. Clinical Outcome Events

Cirrhosis Stage	Prior Decompensating Event	Qualifying Clinical Outcome Events
Compensated	Not applicable	• New onset clinically evident ascites requiring chronic diuretics • Variceal hemorrhage • New onset overt hepatic encephalopathy requiring hospitalization
Decompensated	Ascites, not clinically evident (no ascites or only detectable by ultrasound) on a stable dose of diuretics	• Variceal hemorrhage • Worsening ascites requiring paracentesis • New onset overt hepatic encephalopathy requiring hospitalization
Decompensated	Prior variceal hemorrhage	• Recurrent variceal hemorrhage • New onset clinically evident ascites requiring chronic diuretics • New onset overt hepatic encephalopathy requiring hospitalization
Decompensated	Hepatic encephalopathy, currently with no or mild encephalopathy (Stage 1) on lactulose and/or rifaxmin	• Variceal hemorrhage • New onset clinically evident ascites requiring diuretics • Worsening overt hepatic encephalopathy requiring another hospitalization

Subjects will be evaluated for clinical outcome events at each study visit.

5.3.9 Liver Function by Methacetin Breath Test

The ^{13}C-methacetin breath test (MBT) is a noninvasive tool to assess liver microsomal capacity to metabolize the nonradioactive ^{13}C-labeled methacetin. Details of the MBT are provided in IDN-6556-14 Protocol, Appendix II. The change from baseline in cumulative percentage dose recovery of the metabolized ^{13}C-methacetin 30 minutes after ingestion of the test substrate (cPDR30) will be considered as an exploratory endpoint.

The MBT will be performed at select sites at Screening and Weeks 24 and 48. Sites who elect to participate in the MBT measurements will perform the test at the specified visits for all subjects enrolled at their site.

5.3.10 Liver Stiffness by Transient Elastography (using FibroScan®)

Transient elastography is a non-invasive, reproducible method for measuring liver stiffness that correlates with liver fibrosis and will be measured using FibroScan®. Transient elastography measurements will be performed at select sites at Screening and Week 24 and 48. Sites who elect to participate in the transient elastography measurements will perform the test at the specified visits for all subjects enrolled at their site.

The value for liver stiffness is collected on the eCRF and is the median out of at least 10 successful measurements taken, based on the definition of reliable and representative evaluation of the stiffness of the liver (i.e., success rate of >60% and IQR/median <30%, per IDN-6556-14 Protocol, Appendix III). The change from baseline in liver stiffness at Weeks 24 and 48 will be considered as an exploratory endpoint.

5.4 COMPLIANCE, EXPOSURE, AND SAFETY ENDPOINTS

5.4.1 Treatment Compliance

Study drug dispensation information and treatment compliance data will be collected throughout the study. The number of capsules dispensed, number of capsules returned, number of missed doses and study drug interruptions data will be collected. Treatment compliance will be calculated in the database and is defined as: the number of capsules dispensed minus the number of capsules returned divided by the number of expected doses. Expected doses will be the same for all subjects regardless of when they complete/discontinue study and will be the number of doses to be taken through week 24 and 48, that is 336 doses by Week 24 visit and 672 doses by Week 48.

5.4.2 Extent of Exposure

Overall treatment exposure will be derived as the number of days from the first date of study drug administration to the last date of study drug administration (inclusive of the first and last dates of dosing), regardless of study drug interruptions. Person years of exposure will be derived as the overall exposure (days) divided by 365.25 days.

5.4.3 Adverse Events

Adverse events will be collected during the study and will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). All events with a start date or worsening on or after the first date of study drug administration to the last study visit date are defined as treatment-emergent adverse events (TEAEs).

Adverse events classified as definitely, probably, and possibly related to study drug will be summarized together as related AEs. Adverse events classified as unlikely and not related will be summarized together as not related AEs. For summary purposes, the AE relationship will be summarized as related, not related, or unknown.

Exposure-adjusted AE rate (EAER) is defined as the number of events divided by subject years of exposure. Exposure-adjusted AE rate will be provided for summarizing AEs overall.

5.4.4 Gallbladder Events and Ultrasound

At each study visit, investigator will assess the subject for any symptoms potentially consistent with biliary colic (e.g., sharp right upper quadrant pain, nausea, vomiting, fever [with cholecystitis], etc.), indicated as "yes or no" on the eCRF.

If the subject reports symptoms of potential biliary colic or other symptoms suggestive of gallbladder disease on directed questioning or spontaneously reports such symptoms at any other time during the study, a follow-up liver and gallbladder ultrasound (including assessment of the biliary tree and any gallbladder abnormalities) should be performed.

5.4.5 Liver Monitoring

Liver transaminases (ALT, AST), total bilirubin, cholestatic marker (ALP), INR, and eosinophils will be monitored at each study visit during the study. Given that the patient population to be studied has underlying liver disease and therefore is likely to have abnormalities in liver transaminases and/or bilirubin at baseline, the guidelines for liver monitoring applied in this study incorporate suggestions provided to the Sponsor from the Food and Drug Administration regarding liver monitoring for subjects who have abnormal baseline values of liver transaminases and/or total bilirubin.

In addition, the following categories for ALT, AST, and total bilirubin will be derived for baseline to Week 24, baseline to Week 48, and Week 24 to Week 48.

- Maximum post-baseline comparison to upper limit of the normal range (ULN):
  - ALT
    - ≤ULN
    - >ULN and ≤3× ULN
    - >3× ULN and ≤5× ULN
    - >5× ULN and ≤8× ULN
    - >8× ULN
  - AST
    - ≤ULN
    - >ULN and ≤3× ULN
    - >3× ULN and ≤5× ULN
    - >5× ULN and ≤8× ULN
    - >8× ULN
  - Total bilirubin
    - ≤ULN
    - >ULN and ≤1.5× ULN
    - >1.5× ULN and ≤2× ULN
    - >2× ULN and ≤3× ULN
    - >3× ULN
- Maximum post-baseline value >2× baseline
  - and baseline >ULN

- Maximum post-baseline value >3× baseline
  - and baseline <ULN
- Post-baseline ALT / AST with concurrent (i.e., same post-baseline visit) total bilirubin
  - ALT >3× ULN and concurrent total bilirubin >2× ULN
  - AST >3× ULN and concurrent total bilirubin >2× ULN
  - ALT or AST >3× ULN and concurrent total bilirubin >2× ULN
  - ALT and AST >3× ULN and concurrent total bilirubin >2× ULN
- Possible Hy's Law
  - ALT >3× ULN and not necessarily concurrent (i.e., across all post-baseline visits) total bilirubin >2× ULN
  - AST >3× ULN and not necessarily concurrent total bilirubin >2× ULN
  - ALT or AST >3× ULN and not necessarily concurrent total bilirubin >2× ULN
  - ALT and AST >3× ULN and not necessarily concurrent total bilirubin >2× ULN

5.4.6 Laboratory Measurements

Laboratory tests include measurements of parameters included in hematology and coagulation, chemistry, and metabolic panels. All specific laboratory tests are listed in detail in the protocol. Normal ranges will be provided by the central laboratory and used to derive classifications of laboratory measurements outside of the normal range.

Hematology and coagulation and chemistry panel tests will be collected at each visit. The metabolic panel will be collected on Day 1, Week 24, and Week 48. Changes from baseline will be derived for each applicable laboratory test. Laboratory tests with categorical results will have no derivations made.

5.4.7 Vital Signs and Weight

Vital signs include systolic and diastolic blood pressure, pulse, respiratory rate, and temperature. Body weight and height will be collected at protocol specified visits (weight at all visits, height at screening only), and body mass index (BMI) will be derived as [weight in kilograms (kg)] / [height in meters $(m)^2$].

Changes from baseline at each study visit in vital signs, weight, and BMI will be derived.

5.4.8 Electrocardiograms

A 12-lead ECG will be performed at Screening, Week 4, Week 24, Week 48, and Follow-up. Electrocardiogram data to be collected includes ventricular rate, PR interval, RR interval, QRS interval, QT interval, QTc interval, and findings (i.e., normal or abnormal). Derivations for Bazett's QT correction (QTcB) and Fridericia's QT correction (QTcF) intervals will be conducted in the study database.

Post-dose maximum values in QTcB and QTcF will be classified into the following categories, in milliseconds (ms):

- <450
- >=450 to <480

- >=480 to <500
- ≥500

Changes from baseline in PR interval, QRS interval, QT interval, QTc interval, QTcB interval, and QTcF interval will be derived. Changes from baseline values in QTcB and QTcF will be classified into the following categories, in ms:

- <0
  * >=0 to <30
- >=30 to <60
- ≥60

5.4.9 Medications

All additional medications taken any time during the study will be collected at each study visit. Medications will be classified into 1 of the following 3 types:

- Prior Medication: any medication stopped prior to the first date of study drug administration
- Concomitant Medication: any medication started prior to and stopped on or after the first date of study drug administration
- New Concomitant Medication: any medication started on or after the first date of study drug administration

All medications will be coded using WHODRL.

5.4.10 Physical Examination

A comprehensive physical examination will be performed at the Screening, Week 24, and Week 48 visits including examination of: general appearance, skin, head, eyes, ears, nose, throat, neck (including thyroid), lymph nodes, chest, heart, abdomen (including liver and spleen examination), extremities, and nervous system. A focused physical examination limited to relevant assessments and based on any symptoms or concerns related to a particular body system will be performed at all other visits. Each subject's weight, in street clothes with shoes and outerwear off, will be assessed at each study visit and BMI will be calculated. Assessment of Child-Pugh clinical features will be performed at all visits.

No derivations will be made for physical examination components results.

5.4.11 Alcohol Assessments

The Alcohol Use Disorders Identification Test (AUDIT) is a 10-item questionnaire developed to screen for excessive drinking, helping to identify whether a person has hazardous (or risky) drinking, harmful drinking, or alcohol dependence. The AUDIT-C is a 3-question version of AUDIT that performed similarly to the full AUDIT for detecting heavy drinking and/or active abuse or dependence.

The Skinner Alcohol Dependence scale is a 25-item questionnaire that provides a quantitative measure of the severity of alcohol dependence with scores in the 1st, 2nd, 3rd, and 4th quartiles

corresponding to low, intermediate, substantial, and severe levels of alcohol dependence being likely.

The AUDIT and Skinner Alcohol Dependence scale will be used during Screening to screen for any evidence of alcohol abuse, and the AUDIT-C will be collected at Weeks 12, 24 and 48 to monitor for any evidence of active alcohol abuse.

No derivations of the Skinner Alcohol Dependence scale, AUDIT, or AUDIT-C will be made.

6 ANALYSIS POPULATIONS

The Enrolled Set consists of all subjects who signed an informed consent form.

The randomized set consists of all subjects randomized.

The Full Analysis Set (FAS) consists of all randomized subjects who have received at least one dose of study drug. The analyses based on the FAS population will be conducted on an intention-to-treat principle (i.e., all subjects will be analyzed with the group to which they were randomly assigned).

The Per Protocol Set (PPS) consists of all subjects in the FAS population who have a post-baseline HVPG measurement and do not have any significant protocol deviations. Designation into the PPS will be determined by a blinded data review prior to database lock.

The Safety Analysis Set consists of all randomized subjects who have received at least one dose of study drug on an "as treated" basis (i.e., all subjects will be analyzed by the treatment they have taken).

7 GENERAL STATISTICAL CONSIDERATIONS

There will be two analyses conducted for this study, a primary analysis and a final analysis. The primary analysis to be conducted after all randomized subjects have either discontinued prior to week 24 or completed the Week 24 planned study visit. This primary analysis will be considered as a database freeze and will include all data collected at the time of the database freeze (i.e., will include data collected after the Week 24 study visit). The final analysis will include all subjects and all completed study visits. The clinical study report will be written after the final analysis is conducted.

Analyses will be conducted using SASv9.3 or higher and pooled across all enrolling sites. Data will be summarized descriptively with categorical variables being summarized using number of observations and percentages, and continuous variables being summarized by n, mean, median, standard deviation, minimum, and maximum, unless otherwise specified.

All disposition, demographics, and baseline summaries will be based on the Enrolled set, Randomized set and FAS. The efficacy analyses will be based on the FAS, with the supportive analysis of the primary endpoint also based on the PP. All safety analyses will be based on the Safety Analysis set.

All summaries by treatment group will include each of the planned 3 treatment groups, along with the 3 emricasan treatment groups combined (i.e., all emricasan) and all subjects combined (i.e., total).

The baseline for the HVPG, MELD, and Child-Pugh score is defined as the last non-missing value observed prior to and including Day 1. All other continuous endpoints will have their baseline defined as the average of all observed values collected prior to and including Day 1. For programming purposes, 7 days will be defined as 1 week.

Formatting of numerical results will include the following:

- Mean values will be reported to 1 decimal place more than the data collected.
- Standard deviation values will be reported to 2 decimal places more than the data collected.
- Median, minimum, and maximum values will be reported to the same decimal place as the data collected.
- Percentages will be reported to 1 decimal place.
- Where confidence intervals (CI) are to be reported, CIs for means will be reported to the same decimal places as mean values; CIs for percentages will be reported to 1 decimal place.
- All p-values will be reported to 3 decimal places.

8 SUMMARY OF STUDY POPULATION DATA

All study population data will be provided in subject specific listings.

8.1 SUBJECT DISPOSITION

Subject disposition will be summarized descriptively by treatment group. The disposition summary will include results for subjects enrolled, subjects randomized, subjects treated, subjects included in each analysis population, discontinuation of study drug, reason for study drug discontinuation, study completion status (at both Week 24 and Week 48), and reason for study withdrawal.

A separate table will also be provided using the enrolled set that summarizes enrollment, randomizations, mis-randomizations, stratifications, and mis-stratifications.

8.2 PROTOCOL DEVIATIONS

Protocol deviations will include the deviation recorded, the deviation category, and significance classification. The deviation category refers to the general aspect of the study the deviation affects. Classifications of protocol deviations will be determined by the clinical trial study team prior to database lock during the data review meeting.

Protocol deviations will be summarized descriptively by treatment group. A listing of protocol deviations by unique subject number will be provided.

8.3 DEMOGRAPHICS, HEIGHT, AND WEIGHT

Subject demographics will be summarized descriptively by treatment group. The demographics summary will include results for subject age, age group, gender, race, ethnicity category, height, weight, and BMI. Age group is defined as the following categories:

- ≥18 - <40 years
- ≥40 - <65 years
- ≥65 years

8.4 BASELINE CHARACTERISTICS

Baseline characteristics include the following:

- Baseline results for HVPG
- Stratum - baseline NSBB use (Yes, No)
- Stratum - baseline compensation status
  - Compensated
  - Decompensated
    - Decompensation type (i.e., ascites, variceal bleed, and/or encephalopathy)
- Varices (none, small, medium, large)
- Compensated status with varices
  - Compensated with none or small varices
  - Compensated with medium or large varices

  - Decompensated with none or small varices
  - Decompensated with medium or large varices
- MELD score, MELD classification (i.e., ≤14, ≥15), and MELD score components (i.e., total bilirubin, serum creatinine, and INR)
- C-P score, C-P classification (Section 5.3.3), and C-P score components (i.e., total bilirubin, serum albumin, ascites, hepatic encephalopathy, and INR)

All baseline characteristics will be summarized descriptively by treatment group. A listing of baseline characteristics will be provided by unique subject number.

8.5 MEDICAL HISTORY

In addition to general medical history, specific medical history conditions to be collected at baseline will include type 2 diabetes mellitus, impaired fasting glucose, impaired glucose tolerance, obesity, dyslipidemia (hypertriglyceridemia, low HDL-C), hypercholesterloaemia (high LDL-C), hypertension, gallstones, biliary colic, cholecystectomy, polycystic ovary syndrome, gout, coronary artery disease, myocardial infarction, angina, and stroke.

The percentage of subjects with these specific medical history conditions will be summarized descriptively by treatment group. A listing of medical history events will be provided by unique subject number.

8.6 ALCOHOL SCALES

The AUDIT and Skinner Alcohol Dependence Scale will be collected at baseline. These scales will be summarized descriptively by treatment group and will include the classification for level of alcohol dependence.

The AUDIT-C will be collected post-baseline and will be summarized descriptively at each visit by treatment group.

9 EFFICACY ANALYSES

All efficacy endpoints will be provided in listings by unique subject numbers and visit.

9.1 PRIMARY EFFICACY ANALYSIS

The observed HVPG at screening, Week 24, and change from baseline to Week 24 will be summarized descriptively by treatment groups. Treatment comparisons for continuous change from baseline in HVPG will be conducted using a fixed effects analysis of covariance model (ANCOVA) using treatment (4 levels), baseline HVPG, baseline compensated/decompensated status, and NSBB use as fixed effects. Least-square adjusted means (LSMeans) and 95% CIs at Week 24 will be provided for each treatment, along with the estimated difference in LSMeans and 95% CIs between each emricasan treatment group and placebo. A one-step Dunnett's test will be applied to adjust for the multiple comparisons of each emricasan treatment group with placebo. Missing data for the primary endpoint will be imputed using a multiple imputation (MI) approach, as described in Section 12, for the primary analysis.

The primary analysis will include all subjects classified into the FAS, and will not exclude subjects for study drug interruptions or discontinuations. The measure of intervention will be HVPG at Week 24 and the magnitude of the treatment effect will be estimated by the LSMeans difference relative to placebo at Week 24.

Sensitivity analyses of the primary endpoint may include other methods for the handling of missing data such as observed cases, log-scale, PPS, and as well as emerging methods (if applicable)

Model-based dose response analysis will be conducted outside the scope of this SAP and results will be summarized in a separate report.

9.2 SECONDARY EFFICACY ANALYSES

All secondary analyses will be conducted using the FAS. No multiplicity adjustments will be made for any secondary analyses.

9.2.1 Secondary Efficacy Analysis: HVPG Response 20% Reduction

The HVPG response 20% reduction will be summarized descriptively by treatment group at baseline and Week 24. Treatment comparisons for HVPG response 20% reduction will be conducted using a fixed effects logistic regression model using treatment (4 levels), baseline HVPG, baseline compensated/decompensated status, and NSBB use as fixed effects. The odds ratios and risk differences between any two treatment groups will be reported with their corresponding asymptotic 95% CIs.

9.2.2 Secondary Efficacy Analysis: Caspase 3/7

Data for caspase 3/7 will be summarized descriptively for raw values, CFB, and pCFB at each visit by treatment group.

Treatment comparisons will be based on the analysis of CFB on log-transformed caspase 3/7 at Week 24 and 48. It will be conducted using a linear mixed-effects model using treatment (4 levels), baseline value, visit, treatment-by-visit interaction, baseline compensated/decompensated status, and NSBB use as fixed effects. The unstructured covariance model will be used. The

LSMeans and 95% CIs at 24 and 48 weeks will be reported for each treatment group, along with the estimated difference in LSMeans and 95% CIs for each pairwise comparison. This model will use all available information for all subjects; hence no imputations for missing data will be implemented.

Results will be back-transformed to provide a between-treatment ratio on the post-to-baseline ratio. If log-transformations do not allow for this parametric model to be conducted, then a non-parametric method will be conducted and documented in the clinical study report.

9.2.3 Secondary Efficacy Analysis: ALT

Data for ALT will be summarized descriptively for raw values, CFB, and pCFB at each visit by treatment group. Treatment comparisons will be based on the analysis of CFB on log-transformed ALT. It will be conducted using the same type of mixed-effects model as described in Section 9.2.2.

9.3 EXPLORATORY EFFICACY ANALYSES

All exploratory analyses will be conducted using the FAS population. No multiplicity adjustments will be made for any exploratory analyses.

9.3.1 Exploratory Efficacy Analysis: HVPG Response 10% Reduction

The exploratory endpoint of HVPG response, defined in Section 5.3.1 above, will be summarized descriptively by treatment group and visit. Treatment comparisons for the HVPG response defined as 10% reduction at Week 24 will be conducted using the same type of fixed-effects model as described in Section 9.2.1.

9.3.2 Exploratory Efficacy Analysis: MELD Score, Regression, and Progression

The observed, CFB, and pCFB in MELD scores will be summarized descriptively by treatment group and visit. Similarly, each definition of MELD regression and progression at Week 24 and 48, defined in Section 5.3.2 above, will be summarized descriptively by treatment group and visit.

Treatment comparisons at Weeks 24 and 48 for change from baseline in MELD score will be conducted using the same type of mixed-effects model as described in Section 9.2.2. This same analysis will also be conducted for the change from baseline in MELD score.

9.3.3 Exploratory Efficacy Analysis: C-P Score and Response

The C-P score and each component will be summarized descriptively in two formats. The first format is descriptively summarizing each category by treatment group and visit. The second format is descriptively summarizing the shift in categories between baseline and Week 24 and 48 by treatment group.

Each definition of regression and progression for C-P score will be summarized descriptively at Weeks 24 and 48.

9.3.4 Exploratory Efficacy Analysis: Biochemical, Fibrosis, and Functional Biomarkers

Biomarker endpoints will be summarized descriptively at each visit by treatment group on raw value, CFB, and pCFB. Treatment comparisons at Weeks 24 and 48 for change from baseline in the raw values and the log-transformed raw values of biomarker endpoints will be conducted using the same type of mixed-effects model as described in Section 9.2.2.

9.3.5 Exploratory Efficacy Analysis: CLDQ and SF-36

The overall CLDQ score, each CLDQ domain score, SF-36 PCS, SF-36 MCS, and each SF-36 subscale score will be summarized descriptively at each visit by treatment group. Change from baseline values for each domain at each post-baseline visit will be summarized descriptively by treatment group and overall. Treatment comparisons at Weeks 24 and 48 for change from baseline in CLDQ and SF-36 scores will be conducted using the same type of mixed-effects model as described in Section 9.2.2.

9.3.6 Exploratory Efficacy Analysis: Clinical Outcome Events (Decompensation)

Classification of a clinical outcome event will be summarized descriptively by treatment group for the treatment period. Treatment comparisons for clinical outcome events through Weeks 24 and 48 will be conducted separately using the same type of fixed-effects model as described in Section 9.2.1.

9.3.7 Exploratory Efficacy Analysis: Liver Metabolic Function (Methacetin Breath Test)

Liver metabolic function will be summarized descriptively at each visit by treatment group. Change from baseline values in liver metabolic function at each post-baseline visit will be summarized descriptively by treatment group and overall. Treatment comparisons at Weeks 24 and 48 for change from baseline in liver metabolic function will be conducted using the same type of mixed-effects model as described in Section 9.2.2.

9.3.8 Exploratory Efficacy Analysis: Liver Stiffness

Liver stiffness will be summarized descriptively at each visit by treatment group. Change from baseline values in liver stiffness at each post-baseline visit will be summarized descriptively by treatment group and overall. Treatment comparisons for change from baseline for liver stiffness will be conducted using the same type of mixed-effects model as described in Section 9.2.2.

9.4 EXAMINATION OF SUBGROUPS FOR EFFICACY ENDPOINTS

The primary endpoint will be summarized descriptively by treatment group and overall for the following subgroups:

- Gender (Female, Male)
- Baseline compensation status (compensated, decompensated)
- NSBB use (Yes, No)
- Baseline varices (none/small, medium/large)
- Baseline MELD score ($\leq$14, $\geq$15)
- Baseline Child-Pugh classification (A, B)

- BMI (<30 kg/m^2, ≥30 kg/m^2)
- Type II Diabetes (Yes, No)
- Investigative Site
- Region (United State, Outside of United States)

Additional subgroup analyses will be conducted as exploratory analyses but will be handled outside of these planned analyses. Ad hoc exploratory analyses conducted outside of the SAP may be reported and discussed in the clinical study report, as appropriate.

10 POPULATION PHARMACOKINETICS (PK)

Blood samples will be collected for population PK assessment. A population analysis of time versus emricasan plasma concentration data will be performed using the nonlinear mixed effects modeling approach. The software NONMEM (UCSF, California, USA) will be used to derive the population mean and variance values for specific PK parameters. Additionally, a relationship between PK parameters (or dose) and efficacy, as well as AEs, will be investigated.

In the PK (and pharmacodynamics, PD) analysis, several covariates will be tested and incorporated into the structural model if shown to significantly improve the model's ability to describe the data. The final PK/PD model for emricasan will be obtained from this "full" model using only the covariate relationships that are thought to result in clinically significant alterations in drug PK and/or PD. This model will include data from this study and also from previously conducted Conatus studies.

Results from these population PK/PD models will be provided in a separate report and all analyses are outside the scope of this SAP.

11 SAFETY ANALYSES

All safety endpoints will be provided in listings by unique subject numbers and visit.

11.1 DOSING AND EXTENT OF EXPOSURE

Study drug capsule counts will be summarized descriptively by treatment group and visit. Treatment compliance and extent of treatment exposure endpoints (i.e., days and person years of exposure) will be summarized descriptively by treatment group.

11.2 ADVERSE EVENTS

An overall summary of TEAEs will be summarized descriptively. The incidence of TEAEs will be summarized descriptively by treatment group and overall for each of the following:

- by system organ class (SOC) and preferred term (including EAER)
- by SOC
- by SOC, preferred term, and severity
- by SOC and preferred term for subjects with TEAEs classified as related to study drug
- by SOC and preferred term with events having incidence of ≥5% in any treatment group
- by SOC and preferred term for subjects with study discontinuation
- by SOC and preferred term for subjects with study discontinuation due to AE
- by SOC and preferred term for subjects with study drug discontinuation
- by SOC and preferred term for subjects with study drug discontinuation due to AE
- by SOC and preferred term for subjects with study drug interruption
- by preferred term (alphabetically)
- by preferred term (decreasing incidence in total emricasan group)
- by preferred term for events classified with severity as moderate (decreasing incidence in total emricasan group)
- by preferred term for events classified with severity as severe (decreasing incidence in total emricasan group)
- by preferred term for events classified with severity as moderate or severe (decreasing incidence in total emricasan group)

The incidence of SAEs for SAEs will be summarized descriptively by treatment group and overall for the following:

- by SOC and preferred term
- by SOC and preferred term for subjects with SAEs classified as related to study drug
- by preferred term (alphabetically)
- by preferred term (decreasing incidence in total emricasan group)

Subject listings by unique subject number and visit will be provided for the listings below:

- All AEs
- Subjects who discontinued from the study due to AE

- Subjects who discontinued from study drug due to AE
- Subjects who had a drug interruption due to AE
- Subjects with SAEs
- Subjects with an event of death

11.3 GALLBLADDER EVENTS AND ULTRASOUND

Baseline gallbladder findings on ultrasound will be summarized descriptively by treatment group. The number and percentage of subjects identified as having potential biliary colic symptoms (at least once during the study) based on "biliary colic symptoms" checked yes on eCRF will be summarized descriptively by treatment group. This descriptive summary will also include the number and percentage of subjects with evidence of new gallbladder wall thickening on ultrasound and those with likely gallbladder inflammation events, which will be evaluated by medical review during the blinded data review meeting prior to database lock.

A listing will be provided displaying the baseline medical history related to gallbladder, findings from baseline and follow-up ultrasounds, adverse events from the gastrointestinal SOC, and temperature, white blood cell count, gamma-glutamyl transferase, alkaline phosphatase, total bilirubin, AST, and ALT results during the study for each subject identified as having potential biliary colic symptoms.

11.4 LIVER MONITORING

The baseline value used for purposes of liver monitoring is defined as the mean of all screening and Day 1 values within 70 days of the Day 1 visits, inclusive. Confirmed triggers of liver monitoring as described in Protocol Section 8.1.6 will be summarized descriptively by treatment group.

Subject incidence of the categorizations for ALT, AST, and total bilirubin as described in Section 5.4.5 will be summarized descriptively by treatment group and visit.

An evaluation of drug-induced serious hepatotoxicity (eDISH) plot of all subjects coded by treatment group will be provided, which displays the ALT value (as a multiple of ULN, on a log 10 scale) versus the total bilirubin value (as a multiple of ULN, on a log 10 scale). The eDISH plot will be produced with concurrent results for ALT, AST, and bilirubin by visit and an additional plot that displays the maximum values of ALT and total bilirubin (as a multiple of ULN, both on a log 10 scale).

11.5 LABORATORY MEASUREMENTS

Hematology, coagulation, chemistry, metabolic, and urinalysis laboratory parameters will be summarized descriptively at each visit by treatment group. Where applicable, change from baseline in laboratory tests will also be summarized descriptively at each post-baseline visit by treatment group. Box and Line plots by treatment group will be produced across visits for laboratory parameters of interest.

A shift table will be provided for descriptively summarizing the classification around the normal range of each hematology, coagulation, metabolic, and chemistry laboratory parameter between baseline to each of the minimum and maximum post-baseline values.

An additional shift table will be provided for descriptively summarizing the classification of urinalysis laboratory parameters between baseline and the maximum post-baseline values by treatment group and overall.

Listings will be provided for hematology, coagulation, metabolic, and chemistry laboratory parameters.

11.6 VITAL SIGNS AND WEIGHT

Vital signs, weight, and BMI will be summarized descriptively at each visit by treatment group. Change from baseline in vital signs will also be summarized descriptively at each post-baseline visit by treatment group and overall.

Abnormalities in vital signs will also be summarized descriptively by treatment group and visit.

11.7 ELECTROCARDIOGRAMS

Classifications of observed values and change from baseline values in QTcB and QTcF, as described in Section 5.4.8, will be summarized descriptively by treatment group and visit. Maximum values and maximum change from baseline values during the treatment period will also be summarized descriptively.

11.8 MEDICATIONS

Incidence of prior medications will be summarized descriptively by treatment group, anatomical therapeutic chemical classification (ATC) level, and coded term. Incidence of concomitant, new concomitant, and all concomitant medications will also be summarized descriptively by treatment group, ATC level, and coded term.

11.9 PHYSICAL EXAMINATION

The comprehensive physical examination assessments conducted at baseline, Week 24, and Week 48 will be summarized descriptively at each visit by treatment group. Changes in physical exam status will be summarized descriptively by treatment group using shift tables.

11.10 AUDIT-C

Each alcohol assessment will be summarized descriptively by treatment group and at each visit collected.

12 HANDLING OF MISSING DATA

Missing data for the primary endpoint of change from baseline in HVPG at Week 24 will be imputed using an MI analysis. The amount of missing data will dictate the number of imputed datasets to be derived for the MI analysis.

Assuming an arbitrary missing data pattern, imputed values will be dependent on examination of potential risk factors, such as age, gender, compensated/decompensated status, use of NSBB, and baseline HVPG score. An arbitrary data missing pattern will be assumed and a fully conditional specification regression method will be used to impute missing CFB at Week 24 HVPG values.

The SAS MI procedure will be used to generate multiply imputed datasets with each multiply imputed dataset being analyzed separately as described in the primary analysis. The results for each imputed dataset analysis will then be summarized using the SAS MIANALYZE procedure with the appropriate statistics to be reported.

There will be no imputations made for missing secondary or exploratory endpoints

13 DATA MONITORING COMMITTEE

An independent Data Monitoring Committee (DMC) will review unblinded safety data from this study at periodic intervals. Members of the DMC will not be allowed to participate as Investigators in this study and will not otherwise consult for the Sponsor.

A charter, which will include a detailed description of the scope and the extent of the DMC responsibilities and procedures, will be implemented prior to any data review. These DMC documents (i.e., charter, open and closed meeting minutes, etc.) will be considered part of the study documentation, but not part of the study protocol. The DMC will review data within its general remit to oversee subject safety in the study, and provide recommendations and guidance to the Sponsor in accordance with the procedures stated in its charter.

As part of the remit of the Data Monitoring Committee, an independent Liver Adjudication Committee will assess in a blinded manner the cases of all subjects who:

- Have confirmed elevations in liver parameters that meet the most recent protocol-defined criteria for liver monitoring
- Have adverse events of hepatic injury or failure
- Have serious adverse events that map to the Hepatobiliary system organ class
- Hepatocellular carcinoma (HCC)

The Liver Adjudication Committee will provide an opinion as to whether the elevations in liver parameters are likely drug-induced, and whether the elevation was likely related to blinded study drug. With regard to AEs and SAEs, the Committee will provide an opinion as to whether that event was likely related to the underlying liver disease, other co-morbid conditions or to study medication.

Adjudication responses from the Committee will be logged outside of the study database by the third-party adjudication project manager and sponsor. This log will be used to provide a summary table and listing of subjects reviewed by the Committee and associated outcomes.

All Investigators, responsible IRB/IECs, and applicable regulatory agencies will be informed of any decisions made by the Sponsor based on recommendations from the DMC relating to subject safety that affect the conduct of this study. The Investigators will inform the subjects of such actions, and the protocol and informed consent form will be revised, as appropriate.

14 SUMMARY OF CHANGES TO PLANNED PROTOCOL ANALYSIS

14.1 CHANGES FROM PROTOCOL TO SAP V1.0

The planned treatment pairwise comparisons adjustment method planned for secondary and exploratory endpoints will not be applied as part of this SAP.

14.2 CHANGES FROM SAP V1.0 TO SAP V2.0

- Clarifications made to text throughout the SAP to help clarify further details
- Analysis windows defined for programming purposes of applying unscheduled visit data when planned visit data are missing
- Updated definition of MELD score
- Percent change from baseline variables defined for secondary and exploratory endpoints
- Clarified liver stiffness data to be used based on reliable criteria
- Removed dose response analysis from SAP
- Included treatment comparisons for exploratory endpoints
- Added a diabetes subgroup for efficacy
- Included additional AE summary tables and listings
- Added plots for laboratory parameters of interest
- Clarified details regarding MI analysis

15 REFERENCES

D'Amico G, Garcia-Pagan JC, Luca A, and Bosch J. (2006), Hepatic Vein Pressure Gradient Reduction and Prevention of Variceal Bleeding in Cirrhosis: A Systematic Review. Gastroenterology, 131:1611-1624.

Westfall P, Tobias R, and Wolfinger R. Multiple Comparisons and Multiple Testing Using SAS. 2003, SAS Institute.

Signature Page for CTS-S-0366 v2.0

Approved	James Robinson Document Owner 10-Jul-2018 23:00:38 GMT+0000

Approved	David Hagerty Clinical Operations 10-Jul-2018 23:04:40 GMT+0000

Approved	Jean Chan Monitoring 10-Jul-2018 23:11:25 GMT+0000

Approved	Dan Ripley Project Management 12-Jul-2018 17:03:01 GMT+0000

Approved	Edward Smith Regulatory 13-Jul-2018 17:52:54 GMT+0000

Signature Page for CTS-S-0366 v2.0