STATISTICAL ANALYSIS PLAN (SAP)

Title:	TRANSFORM: A 24-week, Randomized, Placebo-controlled, Double-blind, Phase 2b Trial of Setanaxib in Patients with Primary Biliary Cholangitis (PBC) and Elevated Liver Stiffness
NCT number:	NCT05014672
Unique Protocol ID:	GSN000350
Document Date:	27 June 2024

Statistical Analysis Plan (SAP)

Protocol Title:	TRANSFORM: A 24-week, Randomized, Placebo-controlled, Double-blind, Phase 2b Trial of Setanaxib in Patients with Primary Biliary Cholangitis (PBC) and Elevated Liver Stiffness
Protocol Version No./Date:	1.0/ 06-May-2021 2.0/ 16-Dec-2021 3.0/ 07-Apr-2022 4.0/ 09-Feb-2023 5.0/ 27-Sep-2023
CRF Version No./Date:	6.0/ 03-Aug-2023
SAP Version No./Date:	4.0/ 27 Jun 2024

1.0 Approvals

Sponsor	
Sponsor Name:	Calliditas Therapeutics Suisse SA
Representative/ Title:	
Signature /Date:	
ICON	
Biostatistician / Title:	
Signature /Date:	

2.0 Change History

Version/Date	Change Log
0.1/17 August 2021	Created as new
0.2/20 December 2021	Updated for Protocol Version 2.0 and Sponsor Comments on V0.1
0.3/10 June 2022	Updated to align with TFL shells
0.4/3 October 2022	Updated per Sponsor comments on V0.3 and FDA feedback on V0.2
0.5/15 February 2023	Updated per Sponsor comments on V0.4 and Protocol Amendment version 4.0
1.0/17 February 2023	Updated per Sponsor comments on V0.5 and moving to stable version.
1.1/27 October 2023	Updated per Protocol Amendment version 5.0
1.2/01 November 2023	Updated per Sponsor comments on V1.1
2.0/02 November 2023	Moving to stable version V2.0 from V1.2
3.0/06 June 2024	Updated per Sponsor comments on dry-run #1
3.0/27 June 2024	Updated per Sponsor comments on dry-run #2

3.0 Table of Contents

4.0 Purpose

The Statistical Analysis Plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under Calliditas Therapeutics Suisse SA Protocol GSN000350 Amendment version 5.0.

5.0 Scope

This Statistical Analysis Plan outlines the following:

- Study Objectives
- Study Design
- Study Estimands
- Analysis Sets
- Conventions and Definitions
- Applicable Study Definitions
- Statistical Methods

6.0 Introduction

This SAP should be read in conjunction with the study protocol and case report forms (CRF). This version of the plan has been developed by ICON using the protocol version 5.0 dated 27-Sep-2023 and CRF version 6.0 dated 03-Aug-2023. Any further changes to the protocol or CRF that have an impact on the design, or the planned statistical analysis will require updates to the SAP.

The TRANSFORM study was initiated as a randomized, placebo controlled, double-blind, adaptive Phase 2b/3 study with a 52-week treatment period and a 52-week open-label extension period. Enrolment into the study has been significantly slower than predicted and the feasibility of conducting the study within an acceptable timeframe has proven challenging. This is considered related to new and more frequently used treatment options, such as obeticholic acid (OCA) and off-label use of fibrates. One consequence of this is that fewer subjects are meeting the inclusion criteria for alkaline phosphatase (ALP) ≥ 1.67*ULN.

Considering the challenges, under protocol version 5.0, the study has been revised to a randomized, placebo-controlled, double-blind, **Phase 2b** study of setanaxib in **60-70** subjects with PBC and elevated liver stiffness, to assess the safety and efficacy of the 1200 mg and 1600 mg doses of setanaxib versus placebo over a **24-week treatment period**. The primary endpoint has been amended to evaluate change in ALP (%) on a continuous scale at 24 weeks, rather than the originally planned (binary) composite endpoint of change in ALP and total bilirubin at 52 weeks which requires a significantly larger sample size. The secondary endpoints of change in fatigue and liver stiffness are retained, as is the composite endpoint of changes in markers of cholestasis. The rationale for a 52-week Extension Phase Treatment Period and for an interim analysis is no longer applicable; therefore, the Extension Period, the 12-week Follow-up Period and interim analysis have been removed. A detailed rationale for changes to the study design is described in Protocol Amendment version 5.0 Section 1.2. The results from this Phase 2b study will inform the future development of setanaxib in PBC.

This SAP describes the statistical methods that will be used for the analysis of this Phase 2b study and for summarizing data captured under previous versions of the protocol (v1.0 to 4.0). The first version of the SAP was developed and submitted to the FDA in December 2021, prior to enrollment of the first subject. Since this time, the Investigators, Subjects in the study, Sponsor and ICON have remained blind to randomized treatment assignment. Furthermore, team members operating in a blinded capacity have not been granted access to data concerning ALP and bilirubin values over the course of the study. Consequently, any modifications to the ongoing study have not been influenced in any manner by emerging data or results derived from the study.

The final version of the SAP will be finalized before the database is locked; subsequent to this stage, team members working in a blinded capacity will undergo an unblinding process to reveal the actual randomized assignments.

6.1 Changes from Protocol

- Protocol Amendment version 5.0 Section 8.2: the definition of PPS is more clearly defined as all subjects in the FAS who do not have any important protocol deviations during the 24-week treatment period that would potentially affect the efficacy of the study drug.
- Protocol Amendment version 5.0 Section 8.4.2 states that “Abnormal laboratory values will be listed by subject and summaries of the incidence and frequency by treatment group, scheduled visit, severity, and relationship to IMP will be presented.” Clinically significant abnormalities will be recorded as an adverse event (AE) and the severity/relationship to investigational medicinal product (IMP) will be recorded on the AE CRF page. However specific summaries of other abnormal laboratory values will not contain the assessment of relationship to the IMP or severity as this will not be collected in the laboratory database.

7.0 Study Objectives

7.1 Primary Objective

- To evaluate the effect of setanaxib on ALP at Week 24 in patients with PBC and with elevated liver stiffness and intolerance or inadequate response to ursodeoxycholic acid (UDCA)

7.2 Secondary Objectives

- To evaluate the effect of setanaxib on fatigue at Week 24
- To evaluate the effect of setanaxib on liver stiffness at Week 24
- To evaluate the effect of setanaxib on ALP, fatigue and liver stiffness at Week 24, where setanaxib doses are combined
- To evaluate the effect of setanaxib on pruritus at Week 24
- To evaluate the effect of setanaxib on markers of cholestasis at Week 24
- To evaluate the safety and tolerability of setanaxib over a 24-week treatment period

7.3 Exploratory Objectives

- To evaluate the effect of setanaxib over a 24-week treatment period on markers of liver fibrosis, cholestasis, bile acid metabolism, and liver function
- To evaluate the effect of setanaxib on adverse clinical outcomes
- To further evaluate the effect of setanaxib on patient-reported outcomes
- To determine the pre-dose plasma levels of setanaxib and its metabolite GKT138184

8.0 Study Design

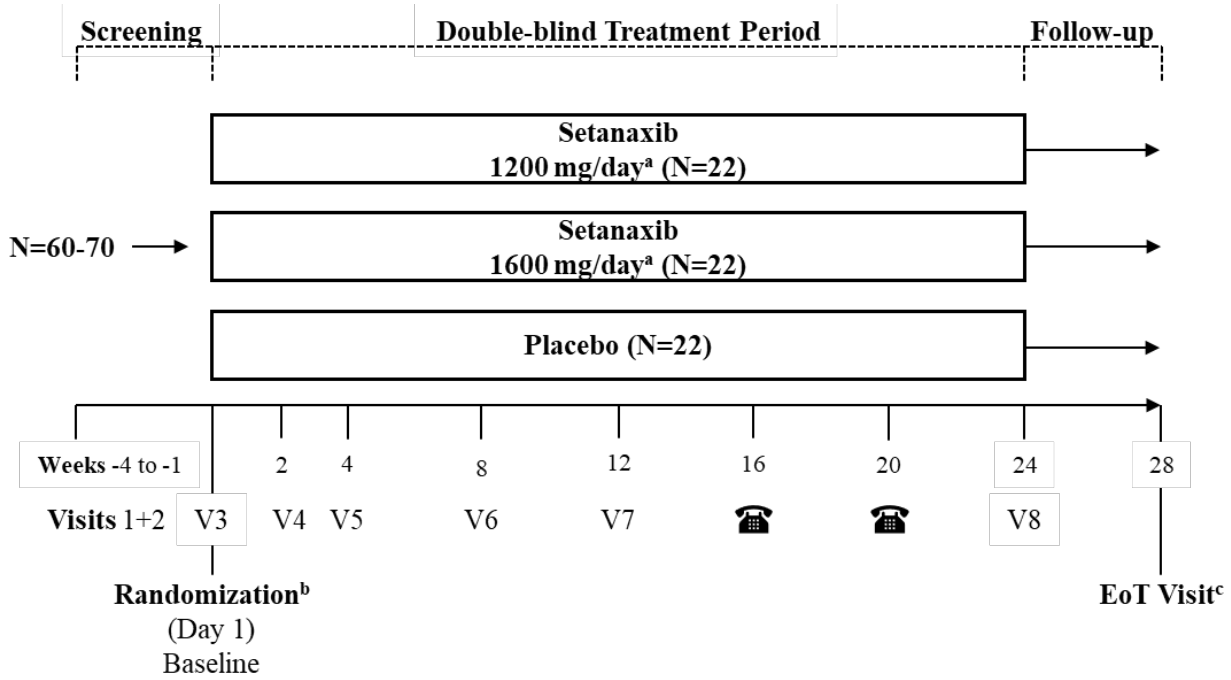
This revised study is a randomized, placebo-controlled, double-blind, parallel-group, multicenter, Phase 2b study assessing oral setanaxib administered as an add-on therapy in subjects with PBC, elevated liver stiffness, and intolerance or inadequate response to UDCA. Intolerance to UDCA is defined as subjects unable to tolerate the full-labelled dose of UDCA in PBC (13-15 mg/kg) due to frequently reported gastrointestinal symptoms such as diarrhoea and abdominal pain. The safety and efficacy of 1200 mg/day and 1600 mg/day setanaxib will be assessed against matching placebo over up to 24 weeks of treatment. It is planned to enroll 60-70 subjects at approximately 80 to 130 investigational centers in North America, Europe, Israel, Australia, and New Zealand.

Subjects enrolled under previous versions (1.0 to 4.0) of the protocol were eligible to receive 52 weeks of treatment in the randomized “Main Part” of the study and then consented to a further 52 weeks of treatment

with setanaxib in the open-label "Extension Phase". The SAP will describe how additional data already collected beyond the 24-week treatment period will be summarized.

The study design is outlined in **Figure 1**. For the schedule of assessments, refer to Table 1 in the Protocol Amendment version 5.0.

Figure 1: Study Design



EoT=End-of-Treatment; V=Visit

[a] The Independent data monitoring committee (IDMC) may recommend change(s) to the setanaxib dose regimen(s), or interruption or discontinuation of an active treatment group(s) based on the regular IDMC safety data reviews (as defined in the IDMC Charter).

[b] Eligible subjects will be randomized to oral setanaxib 1200 mg/day, 1600 mg/day, or placebo, according to a 1:1:1 randomization ratio.

[c] Subjects who permanently discontinue IMP prior to Week 24 Visit (Visit 8) will have an EoT Visit 30 days after the last IMP dose, after which every possible effort should be made to complete the Week 24 Visit (Visit 8).

The study will consist of a 4-week Screening Period, a 24-week Double-blind Treatment Period, and a 4-week Follow-up Period. Under protocol version 5.0 (or later), the total duration of the study for subjects remaining in the study until their final follow-up assessment (end of treatment [EoT] visit 30 days after last investigational medicinal dose [IMP]) will be approximately 32 weeks (approximately 8 months). No interim analyses will be conducted. Unblinded safety data will be summarized by the unblinded programming team for the IDMC, for the purposes of ongoing safety review.

Eligible subjects will be randomized to oral setanaxib 1200 mg/day, 1600 mg/day, or placebo, according to a 1:1:1 randomization ratio, stratified by Screening serum ALP < or ≥3.0×ULN. Refer to Section 8.2 for further details on the randomization in this study.

Baseline assessments will be performed on Day 1 (Visit 3). Post-baseline assessments will be performed at Weeks 2 (Visit 4), 4 (Visit 5), 8 (Visit 6), 12 (Visit 7), and 24 (Visit 8). In addition, study subjects will be contacted by phone by the study site staff at Weeks 16 and 20.

Following permanent IMP discontinuation at any time during any part of the study, subjects will undergo an EoT Visit 30 days after their last dose. If IMP is permanently discontinued the study assessments should

continue until completion of the Week 24 Visit (Visit 8). If it is not feasible to complete all the visits, every possible effort should be made to complete the Week 24 Visit.

The Investigator, the site personnel, the Sponsor and their representatives involved in monitoring and conducting the study, and the subjects will be blinded to treatment assignments until the database is locked and the study is unblinded for analysis.

The IDMC will oversee the safety of participating subjects (see Protocol Amendment version 5.0 Section 9.4.2). The role and responsibilities of the IDMC have been outlined in an IDMC Charter. An Adjudication Committee, who will remain blinded to subject treatment assignment, will adjudicate events requiring IMP interruption as per protocol and the adverse clinical outcomes (see Protocol Amendment version 5.0 Section 9.4.1). The roles and responsibilities of the Adjudication Committee are outlined in an Adjudication Charter.

8.1 Sample Size Considerations

The primary efficacy endpoint is the change in ALP (%) over 24 weeks. It is planned to enroll approximately 60 to 70 subjects. Subjects will be randomized and allocated to placebo, setanaxib 1200 mg/day, or setanaxib 1600 mg/day according to a 1:1:1 randomization ratio.

This study is designed to detect a >25% reduction in ALP in setanaxib-treated subjects versus a 2.5% reduction in the placebo arm with an overall 2-sided $p<0.05$. Standard deviations are assumed to be 19% and 24%, respectively, based on the Phase 2a study with setanaxib and the Phase 3 POISE study with Ocaliva. With approximately 16 evaluable subjects per arm (48 subjects overall) and using a Hochberg step-up procedure (Benjamin and Hochberg, 1995; Section 12.5.1 for details) to control alpha across 2 dose comparisons versus placebo, this study has 88.2% global power to detect at least 1 treatment arm as significantly different from placebo and 68.3% power to detect both treatment arms as statistically significant. Also, for the secondary endpoint of change in fatigue, there will be >80% power to detect a 20% (SD 20%) reduction in fatigue versus a 2.5% (23%) increase in the placebo arm.

An assumed dropout rate of approximately 25% leads to a target sample size of 66 enrolled subjects (22 per arm). A range of approximately 60 to 70 enrolled subjects allows for variability in the assumed dropout rate.

8.2 Randomization

The randomization schema remains unaffected by Protocol Amendment version 5.0. Eligible subjects enrolled according to Protocol Amendment version 5.0 will be randomly assigned to oral setanaxib 1200 mg/day, 1600 mg/day, or placebo, according to a 1:1:1 randomization ratio, stratified by Screening serum ALP < or ≥3.0×ULN.

Randomization will be performed via a centralized Interactive Web Response System (IXRS) contracted through Almac. At registration, the IXRS will assign a unique subject identification number that will be used on all of the subject's eCRFs and serious adverse event (SAE) report forms. If a subject is rescreened (see Protocol Amendment version 5.0 Section 4.3), the rescreened subject should be assigned a new unique screening number.

Any subjects who entered the Extension Phase under protocol versions 1.0 to 4.0 will have received the same setanaxib treatment dose they received during the first 52 weeks of study treatment (the Main Part). Subjects initially randomized to placebo will have been switched to setanaxib 1200 mg/day or 1600 mg/day according to a 1:1 randomization ratio in the Extension Phase. Randomization in the Extension Phase was not stratified. All subjects will have had a new randomization number assigned on enrollment into the Extension Phase, even if they were not re-randomized to a new treatment arm; this is to protect the blind for subjects prior to the completion of the double- part of the study.

During the course of the study, the statistician and the programmers supporting the IDMC analysis will be unblinded and will remain separate from the wider project team throughout the life of the study. The randomization will be performed by the external IXRS provider, Almac. The blinded statistician supporting the study will have no access to the unblinded randomization schedule. The Investigator, the site personnel,

the Sponsor and their representatives involved in monitoring and conducting the study, the Adjudication Committee, and the subjects will be blinded to treatment assignments until the study is unblinded for the primary analysis. For full details refer to the Data Blinding and Documentation plan.

8.3 Types of Planned Analysis

8.3.1 IDMC Analysis

The IDMC will oversee the safety of participating subjects. The IDMC will regularly review an aggregated list of events adjudicated by the Adjudication Committee. The IDMC will meet regularly over the life of the study for data review meetings as specified in the IDMC Charter. The IDMC may recommend change(s) to the setanaxib dose regimen(s), or interruption or discontinuation of an active treatment arm(s) based on the regular IDMC safety data reviews.

During the study there will be at least 2 planned data review meetings. The first IDMC data review meeting took place after approximately 20 subjects have been randomized and the second review meeting will take place after approximately 50 subjects have been randomized. The frequency or number of meetings can be modified, as necessary, by the Sponsor and/or the IDMC. Study enrollment will not be halted during planned IDMC review of safety data.

8.3.2 Interim Analysis (Not Applicable)

The interim analysis is no longer applicable for this revised Phase 2b study.

8.3.3 Primary Analysis

The data cut-off for the primary analysis will occur when 60-70 subjects randomized have had the opportunity to complete or discontinue from the 24-week double-blind treatment period.

The datasets extracted and cleaned for analysis will include all safety and efficacy data collected.

- The datasets prepared for primary analysis will encompass all safety and efficacy data gathered during this period.
- Data collected from 24-weeks to 52 weeks in the Main Part under protocol versions 1.0 to 4.0 will be descriptively summarized.
- Data collected during the extension phase under protocol versions 1.0 to 4.0 will be presented in listings only.

Once all outstanding data queries have been resolved or agreed as unresolvable after data cut-off and the analysis sets agreed, a soft lock of the database will occur when all week 24 data are available, and the primary analysis will be performed. Once all subjects have completed their end of treatment visit, a final hard lock of the database will occur, and no further updates can be made within the electronic data capture system.

9.0 Study Endpoints

9.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change in ALP (%) at Week 24 compared to Baseline.

9.2 Secondary Efficacy Endpoints

The secondary endpoints are:

- Change in fatigue at Week 24 compared to Baseline, as assessed by
    - The Patient-Reported Outcomes Measurement Information System (PROMIS)-short form-Fatigue 7b Daily

  - the Patient's Global Impression of Severity (PGIS) fatigue
  - the Patient's Global Impression of Change (PGIC) fatigue
  - the PBC-40 questionnaire (PBC-40) fatigue domain
- Change in liver stiffness at Week 24 compared to Screening, as assessed by transient elastography (FibroScan®)
- Change in ALP at Week 24 compared to Baseline, where setanaxib doses are combined
- Change in fatigue at Week 24 compared to Baseline, as assessed by the PROMIS short form-Fatigue 7b Daily, the PGIS fatigue, the PGIC fatigue, and the PBC-40 fatigue domain, where setanaxib doses are combined
- Change in liver stiffness at Week 24 compared to Screening, as assessed by transient elastography (FibroScan®), where setanaxib doses are combined
- Change in pruritus at Week 24 compared to Baseline, as assessed by the Worst Itch Numerical Rating Scale (WI-NRS), the PBC-40 itch domain, and the PGIS and the PGIC pruritus
- Changes in markers of cholestasis as assessed by proportion of subjects at Week 24 with:
  - ALP reduction to <1.67×ULN and total bilirubin ≤1xULN and a ≥15% or ≥30% or ≥40% or ≥70% ALP reduction from Baseline, respectively
  - ALP reduction to <1.5×ULN and total bilirubin ≤1xULN and with a ≥40% ALP reduction from Baseline
  - ALP<1xULN and total bilirubin ≤1xULN
  - Total bilirubin <0.6×ULN

9.3 Secondary Safety Endpoints

- AEs. Monitoring for AEs at all Visits
- AEs of special interest (AESIs):
  - Drug-induced liver injury (DILI)
  - Anemia
  - Hypothyroidism
- Laboratory tests:
  - Hematology
  - Biochemistry
  - Urinalysis
  - Thyroid function
- Vital Signs
- 12-lead electrocardiograms (ECGs): clinically significant abnormalities

9.4 Exploratory Endpoints

- Changes in markers of liver fibrosis, as assessed by Enhanced Liver Fibrosis (ELF) score, PRO-C3 (released N-terminal propeptide of Type III collagen), C3M (a peptide of helical collagen Type III degradation), and PRO-C3/C3M ratio at Week 24 compared to Baseline
- Changes in total and conjugated bilirubin over 24 weeks compared to Baseline
- Changes in markers of bile acid metabolism, as assessed by total bile acids, C4 (C4=7α-OH-4-cholesten-3-one), fibroblast growth factor (FGF)19, and FGF21 over 24 weeks compared to Baseline
- Changes in markers of liver function over 24 weeks compared to Baseline, as assessed by:
    - Serum fibrinogen, albumin, and international normalized ratio (INR)
    - Alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transpeptidase (GGT), high sensitivity C-reactive protein (hsCRP), and immunoglobulin M (IgM) indicating liver inflammation and injury
- Assessment over a 24-week period of:
    - All-cause mortality
    - Proportion of subjects requiring a liver transplant (transplant list inclusion)
    - Proportion of subjects with a Model for End Stage Liver Disease (MELD) score of ≥15
    - Proportion of subjects with new onset of variceal/portal hypertension bleed and/or hepatic encephalopathy (West Haven criteria), spontaneous bacterial peritonitis, and ascites requiring treatment
- Change over a 24-week period compared to Baseline in:
    - Social isolation symptoms, as assessed by the PBC-40 social domain
    - Overall health-related quality of life, as assessed by the PROMIS-29 questionnaire
    - EuroQol 5-dimension (EQ-5D) Utility Index
- Predose plasma concentrations of setanaxib and GKT138184 over a 24-week period

Secondary and exploratory endpoints captured for subjects who entered the extension phase under protocol versions 1.0 to 4.0 are described in protocol V4.0.

9.5 Estimand Attributes

The primary and secondary estimands are described in section 12.5.2.

9.6 Analysis Sets

9.6.1 All Enrolled Analysis Set

The All Enrolled analysis set (ENRL) consists of all subjects who signed informed consent for this study. The ENRL set will be used for summaries of disposition data.

9.6.2 Safety Analysis Set

The safety analysis set (SS) will include all randomized subjects who receive at least 1 tablet of IMP or placebo. Subjects included will be analyzed based on actual treatment initially received.

Safety data collected from subjects who entered the extension phase will be listed only.

9.6.3 Full Analysis Sets

The Full Analysis Set (FAS) will include all subjects who receive at least 1 tablet of IMP or placebo during the randomized part and have at least 1 post-baseline ALP value for the primary endpoint. Subjects included will be analyzed based on the treatment group they were randomized to regardless of the actual treatment received.

The FAS will be used for summaries of baseline data, efficacy data, and markers of liver fibrosis. Efficacy data collected from subjects who entered the extension phase will be listed only.

9.6.4 Per Protocol Analysis Set

Per Protocol analysis set (PPS) comprises of all subjects in the FAS who do not have any important protocol deviations during the 24-week treatment period that would potentially affect the efficacy of the study drug. Subjects will be analyzed based on the treatment group they were randomized to regardless of the actual treatment received.

Important protocol deviations are defined in the Protocol Deviation Guidance Document for the study. Prior to database lock and unblinding of the study, a blinded data review of the important protocol deviations (including programmatically determined deviations) will be performed to identify the important deviations impacting on the efficacy endpoints. This analysis set will be defined in a blinded manner prior to unblinding the study and agreed with Sponsor.

Deviations impacting efficacy that may lead to exclusion from the PPS will include:

- Deviations related to inclusion/exclusion criteria
- Incorrect treatment received (i.e. the actual treatment received is different to the initially randomized treatment at any point during the first 24-weeks of treatment)
- Non-compliance with study drug intake (≤80%) during the first 24-weeks of treatment
- Efficacy assessments at Week 24 performed significantly out of the protocol specified window
- Missing baseline data for the primary endpoint
- Use of prohibited concomitant medication that could have impacted the efficacy of study treatment.

9.6.5 PK Analysis Set (PKS)

PKS includes all subjects who receive at least 1 dose of setanaxib and have at least 1 measured concentration at a scheduled PK time point post-dose. Subjects may be excluded if they have an important AE or protocol deviations that may affect PK.

10.0 Conventions and Derivations

10.1 Baseline and Change from Baseline

10.1.1 Baseline Values for Liver Biochemistry

Baseline values for liver biochemistry (**ALP, ALT, AST, total and direct bilirubin, and GGT**) are determined by the geometric mean of the last (up to) 3 measurements preceding the first administration of study drug. This includes unscheduled and screening visits. The formula for the geometric mean (GM) given three measurements a, b, and c is expressed as follows:

$$GM = \exp(\log(a) + \log(b) + \log(c)\ /3)$$

For subjects who have records for both Screening/Visit 1 and Baseline Visit 3, the baseline will be calculated as the geometric mean of these two values. If there is an additional unscheduled pre-dose result, the geometric mean will be calculated using up to three values.

For subjects who have records for Screening/Visit 1, Screening/Visit 2, and Baseline Visit 3, the Screening/Visit 1 result should be disregarded due to the subject's re-screening. The baseline should be calculated as the geometric mean of the Screening/Visit 2 and Baseline Visit 3 results. If there is an additional unscheduled pre-dose result, the geometric mean should include up to three values.

10.1.2 Baseline Values for Patient Reported Outcome Assessments

For the Patient Reported Outcome (PRO) assessments PROMIS-short form-Fatigue 7b Daily, PGIS Fatigue and Pruritus and WI-NRS introduced in protocol version 2.0, subjects enrolled under protocol version 1.0 will not have a baseline value collected. The baseline value for subjects enrolled under version 1.0 for these questionnaires will be imputed prior to analysis (Section 12.5.3). For the PBC-40, only the recall period was changed and the baseline value captured under protocol version 1.0 is considered relevant.

10.1.3 Baseline Values for Other Endpoints

For all other efficacy and safety endpoints, baseline is defined as the last non-missing assessment obtained prior to the first administration of study drug. Where time of assessment is not collected and the assessment is collected on the same date as the first dose of IMP, the assessment will be assumed to be per protocol i.e. prior to first dose administration. If multiple continuous assessments for safety and efficacy endpoints occur on the same day as first dose and no time is recorded, the arithmetic mean will be used as baseline. If multiple categorical assessments occur on the same day as first dose and no time is recorded, the worst severity will be used as the baseline value.

10.1.4 Baseline Values for Extension Phase

For subjects enrolled into the Extension Phase under protocol versions 1.0 to 4.0, “Extension Baseline” value will be defined as the value obtained prior to receiving the first dose of the active treatment. Thus, for subjects who were initially randomized to receive placebo and then switched to setanaxib in the extension phase, the value obtained prior to receiving the first administration of active drug will be the extension baseline value.

10.1.5 Change from Baseline

If the value at baseline and at post-baseline timepoint are not missing, change from baseline at any post-baseline time point will be defined as:

$$Change\ from\ baseline = value\ at\ post\ baseline\ time\ point - value\ at\ baseline$$

Percentage change from baseline at any post-baseline time point will be defined as:

$$Percentage\ change\ from\ baseline = 100 * \frac{(value\ at\ post\ baseline\ time\ point - value\ at\ baseline)}{Value\ at\ baseline}$$

Fold change from baseline at any post-baseline time point will be defined as the below ratio:

$$Fold\ change\ from\ baseline = \frac{value\ at\ post\ baseline\ time\ point}{value\ at\ baseline}$$

For the primary endpoints, reductions in ALP from baseline will be defined as:

$$Reductions\ in\ ALP\ from\ baseline\ = 100 * \frac{(value\ at\ baseline - value\ at\ post\ baseline\ time\ point)}{value\ at\ baseline}$$

10.2 Study Day

Study Day 1 is defined as the first date of study treatment.

- For dates prior to Study Day 1, the Study Day is calculated as:

$$Study\ Day\ = (Date\ of\ Interest/Assessment - Date\ of\ Trial\ Day\ 1)$$

- For dates on or post Study Day 1,

$$Study\ Day\ = (Date\ of\ Interest/Assessment - Date\ of\ Trial\ Day\ 1) + 1$$

If the patient was never treated, the date of randomization will be used in lieu of the first date of study treatment in the above calculations.

For subjects enrolled into the Extension Phase under protocol versions 1.0 to 4.0, study day in the Extension Phase may be defined from the date of first dose of setanaxib in the extension phase.

10.3 Visit Windowing

Subjects may not be able to attend visits on protocol specified days. For the purposes of the analysis, observations will be assigned to analysis windows (Tables 1a and 1b). Table 2 displays the analysis windows to be applied for any subjects enrolled under protocol versions 1.0 to 4.0 and entered in the extension phase.

Tabulations will summarize scheduled time points only. All visits (including unscheduled and EoT visits) will be windowed to the appropriate analysis visit regardless of the visit label entered in the eCRF.

For the efficacy data (including a selection of the liver functions tests, liver stiffness as assessed by fibroscan and QoL),

- If more than one measurement falls within a visit window, the geometric mean (for ALP, ALT, AST, total and direct bilirubin, GGT and liver stiffness) and arithmetic mean (for QoL) of all measurements within the window will be used in the analysis.
- For the same endpoints described above, if data for the week 24 assessment is missing, but an EOT result at week 28 is available, the week 24 result will be substituted by the week 28 result. The week 28 result will also be retained in summary tables.
- For a small number of patients (<5) who had liver stiffness values captured at week 8 without a further assessment at week 12, the week 8 result will be carried forward to week 12 in the analysis. The week 8 result will also be retained in summary tables.
- For QoL assessment collected using a daily diary over multiple days prior to the visits (including WI-NRS and PROMIS-Short Form-Fatigue 7b Daily), the last assessment within the window will be used to define the analysis visit for each assessment. When deriving the baseline, the 7 days prior to the start of treatment will be included. If the questionnaire was incorrectly completed over 8 or more days, the additional records will not be included in the calculation of baseline. For post-

baseline assessments, if there are 8 recorded collected, they will be included in the calculation of arithmetic means.

For all other assessments (including all other liver function tests captured in Table 8 [serum fibrinogen, albumin, hsCRP and IgM]), the value closest to the target day will be selected for use in the analysis. If more than one assessment is equidistant from the target day, the assessment later in time will be selected. If there is more than 1 record on the selected day, the arithmetic mean of the continuous assessments will be applied, or the worst assessment (for categorical parameters) will be selected. If there is an EoT assessment and a scheduled assessment within a time point, the scheduled assessment should always be selected for use in the analysis. For analysis for determination of worst case values, for example shift tables to the worst post-baseline value, all post-baseline values collected during the time period defined by the analysis set, including unscheduled visits, will be considered.

Table 1a: Analysis Windows for Main Part

Analysis Visit	Target Study Day	Lower Limit	Upper Limit
Baseline	1		1 (pre-dose)
Day 1	1 (post-dose)		
Week 2	15	2	22
Week 4	29	23	43
Week 8	57	44	71
Week 12	85	72	127
Week 24	169	128	183
Week 28 - EOT	196	184	225

Table 2b: Analysis Windows for Main Part for ADQS domains (PROMIS-Short Form Fatigue 7b Daily, PBC-40, PROMIS-29, EQ5L, PGIC, PGIS and WI-NRS)

Analysis Visit	Target Study Day	Lower Limit	Upper Limit
Baseline	1		1 (pre-dose)
Day 1	1 (post-dose)		
Week 4	29	2	71
Week 12	85	72	127
Week 24	169	128	183
Week 28 - EOT	196	184	225

Table 3: Analysis Windows for data collected under protocol versions 1.0 to 4.0, including Extension Phase (EP)

Analysis Visit	Target EP Study Day	Lower Limit	Upper Limit
Week 36	253	226	309
Week 52	365	310	460

Extension Baseline	1		1 (pre-dose in Extension)*
Week 68/Week 16 Extension	113	2	155
Week 80/Week 28 Extension	197	156	239
Week 92/Week 40 Extension	281	240	323
Week 104/Week 52 Extension	365	324	407

*for subjects on Placebo in the Main Part. Study day is relative to first dose in the Extension Phase.

10.4 Dates of First and Last Dose of Study Drug

The date of first dose of IMP will be the first dose date recorded on the Study Drug page of the CRF. The date of the first dose of IMP in the Extension Phase, will be the first dose date on the Study Drug Administration (Extension Phase) page of the CRF.

If no date of first dose of IMP is available and the subject was not treated, date of first dose of IMP will be missing and the date of randomization will be used in any calculations requiring treatment start date.

Date of last dose of IMP will be the date collected as the completion/discontinuation date on the End of Treatment page of the CRF. If date of last dose of IMP is not available and subject is lost to follow-up based on End of Treatment CRF page, date of last scheduled visit or date of last telephone contact where a dose is recorded, will be used as the date of last dose of IMP.

Subjects who entered the Extension Phase under protocol versions 1.0 to 4.0 will have no End of Treatment page completed after completion of the randomized part. For these subjects, date of the last dose of IMP in the randomized part will be the day prior to the date of the first dose of IMP in the Extension Phase.

10.5 Actual Treatment

To determine the actual treatment in the study for the purpose of safety analysis, subjects will have actual treatment assigned as the dose the subject first received regardless of any dispensing issues at a later visit. For example, if a subject was randomized to placebo but the subjects first dispensed dose in the study was active treatment, they will be assigned actual treatment as the active arm they first received, even if they were later dispensed placebo at all remaining visits.

For subjects entered in the Extension Phase under protocol versions 1.0.to 4.0, subjects on placebo will have actual arm in the Extension Phase set to the active dose they first receive in the extension phase regardless of the active dose they were randomized to. Subjects in the active arm, will remain assigned the actual treatment in the Extension Phase they first had dispensed.

10.6 Prior and Concomitant Medications

Prior medications will be defined as those medications which started prior to the first administration of study drug.

Concomitant medications will be defined as medications starting prior to but ongoing after first administration of IMP or with a start date on or after first dose of IMP in the randomized part.

Concomitant medications for the Extension Phase will be defined as medications starting prior to but ongoing after first administration of IMP in the Extension Phase or with a start date on or after first dose of IMP in the Extension Phase.

Medications which start prior to the first dose of IMP and are ongoing at the time of first dose of IMP will be summarized as both prior and concomitant medications.

10.7 Adverse Events

10.7.1 Treatment Period

The analysis of safety assessments over the **24-week treatment period** (Week 24 + 30 days) will be conducted:

- For patients enrolled under protocol version 5.0 or later who complete the 24-week treatment period, a Week 28 (EoT) visit will occur 30 days after the date and time of last dose of IMP. All safety data will be analyzed.
- For patients enrolled under protocol versions 1.0 to 4.0, treatment may have continued beyond Week 24. Regardless, the safety data will be cut at Week 28 (Week 24 + 30 days) to ensure a consistent safety follow-up period for all patients.

This approach enables the evaluation of the safety and tolerability of setanaxib over a 24-week treatment period, consistent with the study objectives.

The analysis of safety assessments over the **52-week treatment period** (Week 52 + 30 days) will also be conducted:

- Patients enrolled under protocol version 5.0 will contribute safety data up to Week 28 only. All safety data will be analyzed.
- For patients enrolled under protocol versions 1.0 to 4.0 who complete the 52-week treatment period and do not enter the extension phase, a Week 56 (EoT) visit will occur 30 days after the date and time of last dose of IMP. All safety data will be analyzed.
- For patients enrolled under protocol versions 1.0 to 4.0 who entered the extension phase, the safety data will be cut at Week 56 (Week 52 + 30 days) to ensure a consistent safety follow-up period for all patients.

This approach enables the evaluation of the safety and tolerability of setanaxib over the whole randomized treatment period and provides longer-term safety information. This approach will be applied to other safety endpoints.

Note: For the analysis of study drug exposure and compliance (Section 12.3.3), the last dose of IMP in the 24-week treatment period will be cut at week 24, and the last dose in the 52-week treatment period (for subjects who received more than 24 weeks of treatment) will be cut at week 52. "+ 30 days" will not be considered in the drug exposure and compliance.

10.7.2 Treatment-Emergent Adverse Events (TEAE)

The treatment emergent period will be defined in two ways:

(1) TEAEs over a 24-week treatment period will be defined as AEs that occur, having been absent before the date and time of the first dose of the IMP, or have worsened in severity or seriousness after initiating the IMP until Week 28 (Week 24 + 30 days) in the **24-week treatment period**.

For patients enrolled under protocol versions 1.0 to 4.0, the AEs that occur, or have worsened in severity or seriousness after Week 28 (Week 24 + 30 days) will NOT be taken into consideration in TEAEs over a 24-week treatment period.

(2) TEAEs over a 52-week treatment period will be defined as AEs that occur, having been absent before the date and time of the first dose of the IMP, or have worsened in severity or seriousness after initiating the IMP until Week 56 (Week 52 + 30 days) in the **52-week treatment period**.

For patients enrolled under protocol versions 1.0 to 4.0 who entered the extension phase, the AEs that occur, or have worsened in severity or seriousness after Week 56 (Week 52 + 30 days) will NOT be taken into consideration in TEAEs over a 52-week treatment period.

TEAEs will be assigned to the Extension Phase based on AE start date:

- If the start date of the TEAE is on or after the first dose in the Extension Phase, the TEAE will be assigned to the Extension Phase.

10.7.3 TEAE Leading to Discontinuation of Study

A TEAE will be classified as a TEAE leading to discontinuation of the study if the answer to the question "Did the adverse event cause the subject to be discontinued from the study?" is Yes on the AE CRF page.

10.7.4 TEAE leading to IMP Withdrawal

A TEAE will be classified as

- a TEAE leading to study drug withdrawal if the answer to the question "Action Taken with Study Treatment" on the AE CRF page is "Drug Withdrawn" or "Unknown".

and/or

- a TEAE leading to treatment discontinuation if the answer to question "Reason for Discontinuation" on the End of Treatment CRF page is "Adverse Event".

10.7.5 TEAEs Related to IMP

A TEAE will be classified as related to IMP if the relationship to study treatment is indicated as Related on the AE CRF page. Additionally, if the relationship is missing, then the TEAE will be deemed related.

10.7.6 Duration of AEs

Duration of AEs will be calculated as AE end date – AE start date + 1. AE duration will only be calculated when complete dates for the start and stop date are provided.

10.7.7 AESIs

An AE will be classified as an AESI is the question "Is this event of special interest?" is marked as Yes on the CRF.

10.8 Imputation of Missing Dates

For the purposes of assigning events as on-treatment or medications as prior/concomitant imputation of partial dates will be performed. The rules for imputation for AEs and medications are outlined in Appendix 1 and 2 respectively.

The duration of events and time to onset if applicable will not be calculated if start or stop date is partial or missing. Listings will display the collected date and not the imputed date.

10.9 Adverse Clinical Outcomes

A subject has an adverse clinical outcome if any of the following criteria are met:

- Death
- Transplant list inclusion
- MELD score ≥ 15 (confirmed at repeat testing at least 14 days apart)
- Variceal blead
- Portal hypertension bleed
- Hepatic encephalopathy
- Spontaneous bacterial peritonitis
- Uncontrolled ascites requiring treatment

Subjects who discontinue treatment due to adverse clinical outcomes will be identified via medical review and provided to statistics prior to database lock.

10.10 PBC Medication

The initiation of PBC medication while on the study will be treated as rescue medication in the analysis. All PBC medication usage during the study will be collected on the Prior and Concomitant medication CRF page and the category will be “PBC Treatment”. PBC medication that was ongoing at the time of screening will not be considered as impacting the efficacy analysis. If a subject with ongoing PBC medication at time of screening has a temporary medication interruption during the study and the medication is subsequently resumed, this will not be considered a new PBC medication impacting efficacy. The use of UDCA is required during the study, unless the subject is intolerant to UDCA. Obeticholic acid (OCA), fenofibrate and bezafibrate are also permitted during the study. If a stable dose is maintained throughout the study these may not be considered rescue therapies.

All PBC treatment will be reviewed and assessed for the impact on efficacy. If a subject is identified as having a record that potentially impacts efficacy (i.e., the PBC medication started after randomization), all efficacy data collected following the start date of the medication will be treated as missing data for the analysis of the primary, secondary and exploratory endpoints.

The date of PBC medication initiation is defined as the first date/time of PBC treatment identified as impacting efficacy record in the CRF. For imputation of partial dates refer to Appendix 2. An assessment or event will be defined as prior to PBC medication initiation if the date/time of the assessment or event occurred prior to the date/time of the first PBC medication usage. Otherwise, the event will be considered after PBC medication use. If the assessment of event only has date collected, then prior to PBC medication includes assessments or events that occurred on the day of the first PBC medication usage. Time to PBC medication use will be defined as:

$$Time\ to\ PBC\ Medication\ Use = Start\ Date\ of\ PBC\ Medication - First\ dose\ of\ IMP + 1$$

The identification of PBC medication impacting on efficacy will be conducted through a blinded medical review prior to the primary analysis.

10.11 PROMIS-Short Form-Fatigue 7b Daily

The PROMIS-short form-Fatigue 7b Daily comprises of 7 questions asking the subject to evaluate their fatigue since waking up. Each of the questions on a 5-point scale (range 1-5) with a range of possible values for the score for the daily form from 7 to 35. Higher scores indicate worse functioning for fatigue. This measure will be completed at home by the subject via daily diary over 7 consecutive days prior to the study visits.

Total daily raw scores will be obtained by summing item scores, which are converted into T-scores standardized (mean=50, SD=10). The standardized T-score is reported as the final daily score for each participant. The daily raw scores will be converted into T-scores using the table in Appendix 3 (HealthMeasures, 2019). All questions must be answered in order to produce a valid score using the scoring tables, if the subject does not answer all 7 questions completely, the total daily raw scores are set to missing.

The T-scores for baseline and post-baseline visits will be calculated by averaging the daily T- scores before the visit date (inclusive). The calculation of this average value will proceed only if 4 or more daily T-scores are available; otherwise, the final score will be treated as missing for the corresponding visit. The baseline T score will be defined as the mean of the daily T scores over the last 7 days prior to first treatment administration. For post-baseline assessments only, if data over 8 days is collected, all 8 days will be included in the calculation of the average value.

Approximately 15 subjects will have no baseline value for the PROMIS-short form-Fatigue 7b Daily as they were enrolled under Protocol Version 1.0 and this questionnaire was not added until Protocol Version 2.0. These missing data will not be imputed for the main analysis.

10.12 PBC-40 Item Questionnaire

The PBC-40, is a disease-specific health-related quality of life measure validated for use in subjects with PBC (Jacoby et al 2005). The questionnaire comprises of 40 questions in 6 domains related to fatigue, emotional, social, and cognitive function, general symptoms, and itch. An additional 3 questions are asked relating to the subjects general health and well-being.

Subjects will be asked to select his/her most appropriate response for all 43 questions applying a 1-week recall period under protocol version 2.0 or later. For subjects enrolled under protocol version 1.0, a 4-week recall period will apply. All PBC-40 records collected prior to the date of informed consent for the Protocol Amendment (collect on the eCRF page Protocol Amendment Re-Consent) will be assumed to have a recall period corresponding to version 1.0 (4-week). All records on or following the informed consent date for the Protocol Amendment will be assigned a 1-week recall period.

PBC-40 scores are calculated using subject responses to 40 questions in the questionnaire. There are a total of 6 domains across the 40 questions:

- Symptoms
- Itch
- Fatigue
- Cognitive Function
- Emotional function
- Social Function

Items scores often range from 1 to 5 or 0 to 5, with 0 or 1 representing the lowest impact and 5 representing the highest impact of PBC on quality of life. The individual item scores are summed to obtain a total domain score, high scores represent high

impact and low scores indicate low impact of PBC on the quality of life.

Analysis of the PBC-40 is conducted by domain, and scoring is elucidated in the Table 3 below. Instead of considering a cumulative PBC-40 score, data should be evaluated on a domain-by-domain basis.

- In instances where data are missing from a domain, the entire domain should be excluded if less than 50% of items are completed (≥ 50% of items are missing). The total score of the domains will be imputed using the MCMC method in a Sensitivity Analysis see Table 4.
- Conversely, if more than 50% of responses are present then the median value for the completed items in the domain should be ascribed to the missing item in the main analysis. The total score of domains can be calculated and will not be treated as missing.

The score for the six domains will be derived as per Table 3.

Table 4: PBC-40 domain scoring

Domain (No of questions)	**Response Score**
Symptoms (7)	Range: 6 to 35
Q1	Never=5; Rarely=4; Sometimes=3; Most of the time=2; Always=1
Q2, Q4-Q7	Never=1; Rarely=2; Sometimes=3; Most of the time=4; Always=5
Q3	Never=1; Rarely=2; Sometimes=3; Most of the time=4; Always=5; Did not apply/never drink alcohol=0
Itch (3)	Range: 0 to 15

Domain (No of questions)	Response Score
Q8-Q10	Never=1; Rarely=2; Sometimes=3; Most of the time=4; Always=5; Did not apply/no itch=0
Fatigue (11)	Range: 11 to 55
Q11-Q21	Never=1; Rarely=2; Sometimes=3; Most of the time=4; Always=5
Cognitive Function (6)	Range: 6 to 30
Q22-Q27	Never=1; Rarely=2; Sometimes=3; Most of the time=4; Always=5
Social Function (10)	Range: 8 to 50
Q29, Q31	Not at all=1; A little=2; Somewhat=3; Quite a bit=4; Very Much=5; Does not apply=0
Q32	Not at all=1; A little=2; Somewhat=3; Quite a bit=4; Very Much=5;
Q34-Q39	Strongly agree=5; Agree=4; Neither agree not disagree=3; Disagree=2; Strongly disagree=1
Q40	Strongly agree=1; Agree=2; Neither agree not disagree=3; Disagree=4; Strongly disagree=5
Emotional Function (6)	Range: 3 to 15
Q28, Q30, Q33	Not at all=1; A little=2; Somewhat=3; Quite a bit=4; Very Much=5

10.13 PROMIS-29

The PROMIS-29 Profile (v2.1) comprises of 7 domains (anxiety, depression, fatigue, pain interference, physical function, sleep disturbance, and ability to participate in social roles and activities) and a single pain intensity item (Cella et al 2010). Each of the 7 domains consists of 4 questions on a 5-point scale (range 1-5) with a range of possible values for the score for each domain from 4 to 20. Higher scores for the physical function and ability to participate in social roles and activities represent better functioning in those domains. Higher scores in the anxiety, depression, fatigue, pain interference and sleep disturbance indicate worse functioning in these domains. The single pain intensity item is measured on an 11-point rating scale with 0 representing no pain and 10 representing worst pain imaginable.

Total raw scores for each domain will be obtained by summing item scores for each domain, which are converted into T-scores standardized (mean=50, SD=10). The standardized T-score is reported as the final score for each participant. The raw scores will be converted into T-scores using the table in Appendix 4 (HealthMeasures, 2019).

All questions must be answered in order to produce a valid score using the scoring tables.

10.14 Worst Itch Numerical Rating Scale

The WI-NRS is a daily patient-reported measure of itch intensity using an 11-point scale (where 0=no itch and 10=worst itching imaginable). The WI-NRS will be completed and collected via daily diary over 7 days prior to the visits.

The WI-NRS score for baseline and post-baseline visits will be calculated by averaging the daily WI-NRS scores before the visit date (inclusive). The calculation of this average value will proceed only if 4 or more daily scores are available; otherwise, the final score will be treated as missing for the corresponding visit. The baseline WI-NRS score will be defined as the mean of the daily WI-NRS scores over the last 7 days prior to first treatment administration. For post-baseline assessments only, if data over 8 days is collected, all 8 days will be included in the calculation of the average value.

10.15 PGIS and PGIC Fatigue and Pruritus

PGIS and PGIC-Fatigue and PGIS and PGIC-Pruritus will be recorded prior to each study visit on the day prior to this visit. The PGIS is a global index where subjects will be asked to rate the severity of Fatigue and Pruritus on a 5-point scale within the past 7 days, with choices of 1=none, 2=Mild, 3=Moderate, 4=Severe and 5=Very Severe. Higher scores indicate a worst assessment for fatigue and pruritus.

The PGIC is a 7-point scale reflecting a subject's rating of overall change. Subjects will be asked to rate their overall change with fatigue and pruritus with 1=Very much improved, 2=Much improved, 3=Minimally improved, 4=No change, 5=Worse, 6=Much worse, 7=Very much worse. Higher scores indicate greater severity of fatigue and pruritus, respectively. There is no baseline assessment for PGIC.

No missing data will be imputed

10.16 EuroQoL 5-D Utility Index

The 5-level version of the EQ-5D (EQ-5D-5L) comprises a descriptive system with 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and the EQ VAS (Herdman et al 2011; EuroQoL Research foundation 2019).

Each of the 5 dimensions in the descriptive system has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. This decision results in a 1-digit number that expresses the level selected for that dimension. The digits for the 5 dimensions will be combined into a 5-digit number that describes the subject's health state with 1 representing the best clinical status and 5 the worse.

The EQ VAS is a vertical VAS numbered from 0 to 100, where 0=The worst health you can imagine, and 100=The best health you can imagine. The subject will be asked to mark an X on the scale to indicate how his/her health is today, and to write the number he/she marked on the scale in the box next to the scale.

For EQ-5D, questions not answered will be considered missing and no missing data will be imputed. A single EQ-5D utility index value will be derived which reflects how good or bad a health state is according to the preferences of the general population country/region. This utility index can be computed from the EQ-5D-5L descriptive system with utility scores ranging from -0.281 (worst imaginable health state) to 1 (best imaginable health state). The EQ-5D summary index will be derived by applying a formula that attaches weight to each of the levels in each dimension. Country specific weights will be applied (if available) or US weights (if country specific validated weights are not available).

10.17 Definition of LLOQ

Values below lower limit of quantification (LLOQ)

To include safety assessments that are flagged as being below the lower limit of quantification in calculations they will be imputed by

$$\frac{LLOQ}{2}.$$

To not display a misleading precision of the measurements, all descriptive summary statistics with value below the LLOQ will be set to "< <LLOQ>" in tables where "<LLOQ>" is a placeholder for the numerical value of the LLOQ. The LLOQs for Setanaxib and GKT138184 are **5.00 ng/mL** and **1.00 ng/mL**, respectively.

10.18 Imputation of Missing Data

In general, missing safety data will not be imputed.

For the primary and secondary endpoints (Sections 12.5.2 and 12.5.3) analyzed by Mixed Model Repeated Measures (MMRM) analysis, imputation of missing data will be performed in sensitivity analyses, the methods used are summarized in Table 4 below. Data following usage of PBC-related medication after randomization will be set to missing.

Table 5: Summary of Imputation Methods

Analysis	Endpoint	Imputation Method (Sensitivity Analysis)
Primary Analysis	The change in ALP (%) over 24 weeks	Markov chain Monte Carlo (MCMC) monotone imputation as described in Section 12.5.2.2.1
Secondary Analyses	The change in fatigue assessed by the PROMIS-short form-Fatigue 7b Daily	if the final score for the subject is missing for the corresponding visit, MCMC monotone imputation will be used as described in Section 12.5.2.2.1 analogously.
	The change in fatigue assessed by the PBC40 fatigue domain	If >50% of responses are present in the domain then the median value for the completed items in the domain should be ascribed to the missing item in the **main** analysis. If ≤50% of responses are present in the domain, the total score of the domain will be set to missing and imputed using MCMC monotone imputation as described in Section 12.5.2.2.1 analogously
	The change in liver stiffness over 24 weeks	MCMC monotone imputation as described in Section 12.5.2.2.1 analogously

11.0 Interim Analyses (N/A)

No interim efficacy analyses of the Phase 2b study data will be conducted.

12.0 Statistical Methods

All statistical analyses will use SAS® version 9.4 or higher. Unless otherwise noted, categorical variables will be summarized using counts and percentages. Percentages will be rounded to one decimal place, except 100% will be displayed without any decimal places and percentages will not be displayed for zero counts.

Continuous variables will be summarized using the number of observations (n), mean, Standard Deviation (SD), median, Q1, Q3, minimum and maximum, unless otherwise stated. The minimum and maximum values will be displayed to the same level of precision as the raw data, the mean, median, Q1, and Q3 to a further decimal place and the SD to two additional decimal places. The maximum number of decimal places will be 4, unless otherwise noted. P-values will be presented to 3 decimal places, with values less than 0.001 presented as <0.001. If not stated otherwise, p-values from statistical tests will be 1-sided (type I error 2.5%) and 2-sided 95% CIs will be calculated.

For MMRM analysis, the mean will be presented to two decimal places, and the SD and confidence interval (CI) will be presented to three additional decimal places. P-values will be presented to four decimal places.

In general, all data summaries will be presented by treatment group and total group. Efficacy outputs will not have a total column.

All data collected during this study regarding subject characteristics, efficacy and safety will be listed, unless otherwise specified. Listings will not show imputed data but will present data as reported.

Stratification will be performed based on Screening serum ALP < or ≥ 3.0xULN. If there are differences between the randomization strata and the actual strata, these will be listed.

Subjects who are screen failures may be rescreened at the discretion of the Investigator or Medical Monitor. Rescreen subjects will be analyzed under their rescreen subject identifier and data collected under the screen failure identification will be listed under the rescreen identifier. Screen failures will be excluded from all listings and tables if not otherwise noted.

Missing data will be handled as described in Section 10.17. The FAS will be used to analyze endpoints related to the efficacy objectives, the PP will be used for supplementary analysis related to efficacy

objectives, the SS will be used to analyze safety endpoints and assessments related to safety over two treatment periods, as defined in Section 10.7.1. PK analysis will be presented using the PKS.

For information on the control of overall Type 1 error rate in the primary endpoint refer to Section 12.5.1. Formal statistical testing will be conducted for the primary endpoint of change in ALP, secondary endpoint of fatigue (based on PROMIS short form and PBC40 questionnaires) and change in liver stiffness.

12.1 Subject Disposition

A summary of subject disposition will be provided by treatment group and total for the ENRL set. The number and percentage of subjects in the below categories will be summarized:

Analysis	Comment
• Subjects screened (total n only)	
• Subjects randomized	
• Subjects randomized but not treated	
• Subjects treated	
• Subjects who completed at least 24-weeks treatment	*The subject will be defined as completing treatment if they complete the Week 24 exposure CRF visit (or Week 28 in the absence of Week 24).*
• Subjects who discontinued treatment prior to 24-weeks with reasons for discontinuation	
• Subjects who discontinued study prior to 24 weeks with reasons for discontinuation	*A subject will be defined as completing the Study, if they complete Week 24 CRF visit.*
• Subjects who completed at least 52-weeks treatment (protocol version 1.0 to 4.0)	*The subject will be defined as completing treatment at least 52-weeks if they complete the Week 52 exposure CRF visit.*
• Subjects who discontinued treatment prior to 52-weeks with reasons for discontinuation	
• Subjects who discontinued the study prior to 52 weeks with reasons for discontinuation	*The subject will be defined as discontinued study prior to 52-weeks if they do not complete the Week 52 exposure CRF visit.*

• Subjects enrolled in the extension phase
  ○ Subjects assigned to 1200mg
  ○ Subjects assigned to 1600mg
	CRF Study Drug Administration (Extension Phase)

• Subjects treated in the extension phase
  ○ Subjects treated with 1200mg
  ○ Subjects treated with 1600mg
	CRF Study Drug Administration (Extension Phase)

The flow of subjects through each stage of a trial, from recruitment to analysis, will be summarized using CONSORT (Consolidated Standards of Reporting Trials) diagram.

Screen failure and reasons for screen failure will be listed and summarized using the ENRL set.

Subjects disposition and screening information will be listed. For subjects who entered the extension phase, listings will include the dose received in the randomized phase and dose received in the extension phase together with dates.

Enrollment will be summarized by region (North America, South and Central America, Europe, Asia, Africa), country and site by number enrolled, n(%) randomized and n(%) treated for the ENRL set.

Summaries for the number and percentage of subjects included in each analysis set, a breakdown for the reason for exclusion from each analysis set will be provided in listing. The number of subjects on treatment will be provided by time points of 24-week Double-blind Treatment Period under Protocol v5.0 and 52-week Double-blind Treatment Period under Protocol v1.0 to v4.0.

12.2 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized for the FAS. Descriptive statistics will be provided for:

- Sex (Female, Male, Undifferentiated, Unknown)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other, Not Reported, Unknown)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown)
- Age (years) at informed consent (as collected on the Demographics CRF page),
- Age group (⩾18 and <65 years, ⩾ 65 years)
- Baseline body weight, height, and body mass index (BMI). BMI kg/m^2 will be derived as Weight (in kg)/Height (in m)^2
- Randomized Screening serum ALP strata (<3 or ≥3.0xULN based on value used in randomization)

Descriptive statistics will also be provided for the following baseline characteristics:

- Screening serum ALP
- Baseline serum ALP
- Baseline total bilirubin
- Baseline Anti-SP100 (Positive (incl. Weakly Positive), Negative)
- Baseline Anti-GP210 (Positive (incl. Weakly Positive), Negative)
- Baseline AMA titer (Positive, Negative)
- Liver biopsy performed (Performed/Not performed)
- Baseline liver stiffness score as assessed by transient elastography using FibroScan®
- Baseline Child-Pugh Score and Class (Class A, Class B, Class C)
- Baseline MELD score
- Baseline MELD score (proportion of subjects with a score <15 and ≥ 15)

A listing for demographics and baseline disease characteristics will be provided for all subjects in the FAS. Discrepancies between randomization stratification information and actual strata based on the collected Screening serum ALP value will be flagged.

12.3 Treatments

12.3.1 Medical and PBC Treatment History

Medical history conditions will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 24.0 or later. The number and percentage of subjects with each medical history term will be listed

and summarized by the MedDRA system organ class (SOC) and preferred term (PT) using the FAS. A listing for medical history will be presented for the FAS.

PBC treatment history will also be presented for the FAS with the following baseline disease characteristics summarized:

Analysis (Yes/No)	Comment
• Any Prior PBC Treatment prior to or at randomization	*The number and percentage of subjects have at least one ongoing PBC record at randomization, or a record on the PBC Treatment History page of the CRF*
○ UDCA use prior to or at randomization	*Yes, if subjects have an ongoing UDCA record at randomization or a record for UDCA on the PBC Treatment History page of the CRF*
○ OCA use prior to or at randomization	*Yes, if subjects have an ongoing OCA record at randomization or a record for OCA on the PBC Treatment History page of the CRF*
○ Fibrate use (bezafibrate or fenofibrate) prior to or at randomization	*Yes, if subjects have an ongoing fibrate record at randomization or a record for OCA on the PBC Treatment History page of the CRF*
○ Bezafibrate use prior to or at randomization	*Yes, if subjects have ongoing Bezafibrate record at randomization or a record for Bezafibrate on the PBC Treatment History page of the CRF*
○ Fenofibrate use prior to or at randomization	*Yes, if subjects have ongoing* Fenofibrate *record at randomization or a record for* Fenofibrate *on the PBC Treatment History page of the CRF*
• Any Prior PBC Treatment at randomization	*The number and percentage of subjects have at least one ongoing PBC record at randomization.*
○ UDCA use at randomization	*Yes, if subjects have an ongoing UDCA record at randomization*
○ OCA use at randomization	*Yes, if subjects have ongoing OCA record at randomization*
○ Fibrate use (bezafibrate or fenofibrate) at randomization	*Yes, if subjects have ongoing Fibrate record at randomization*
○ Bezafibrate use at randomization	*Yes, if subjects have ongoing Bezafibrate record at randomization*
○ Fenofibrate use at randomization	*Yes, if subjects have ongoing Fenofibrate record at randomization*
• Intolerance to UDCA	*Yes, if a subject has a record on the PBC Treatment History page of the CRF with drug UDCA and reason for Reason Treatment was Discontinued is intolerance, otherwise if subject has an ongoing record for UDCA set to No*

All PBC treatment history will be listed for the FAS, Subjects who started PBC medication after randomization and subjects who stopped PBC medication prior to stopping randomized treatment will be flagged and a further flag if they were assessed to impact efficacy. Time to medication use will be derived and included in the listing.

12.3.2 Prior and Concomitant Medications

Prior Non-PBC medications, concomitant Non-PBC medications (by treatment periods) and concomitant PBC medications (by treatment periods) received with IMP, categorized by medication group according to the World Health Organization Drug Dictionary (WHODRUG) Global version March 2021 B3 or later will be summarized using the SS.

Prior and concomitant medications will be summarized using Anatomic Therapeutic Chemical (ATC) classification and PT for the SS. Where multiple ATC codes are assigned to the same treatment, the blinded Calliditas medical monitor will identify the primary code for inclusion in the summary tables. The number and percentage of subjects using each medication will be displayed together with the number and percentage of subjects using at least one medication within each medication group and subgroup. The summaries will be repeated for each treatment period, as defined in Section 10.7.1.

All Non-PBC medications will be provided in listing and flags will be provided, to indicate prior or concomitant medications, and to indicate treatment periods.

During the blinded data review meeting that was conducted prior to database lock, it was identified that a small number of concomitant Non-PBC medications were included on the concomitant PBC medication eCRF form. A list of associated ATC codes were provided by the Calliditas medical monitor to the ICON programming group to facilitate a move of these medications to the correct summary table.

12.3.3 Extent of Study Drug Exposure

The extent of exposure will be characterized for the SS according to the number of subjects who received at least one dose, the total number of doses received, the duration of therapy (weeks), duration of interruptions, the reasons for interruptions and study drug compliance. The summaries will be repeated for each treatment period, as defined in Section 10.7.1.

At each dispensing visit, subjects will be dispensed kit(s) comprising of 4 bottles containing 70 tablets each. Subjects are expected to take 4 tablets per day, 1 tablet from each bottle, 2 in the morning, 2 tablets in the evening.

The duration of exposure, regardless of temporary interruptions, will be defined as the

$$Duration\ (weeks) = \frac{(Last\ dose\ of\ IMP - First\ dose\ of\ IMP) + 1}{7}$$

The date of last dose of IMP in the 24-week treatment period and date of last dose in the 52-week treatment period (for subjects who received more than 24 weeks of treatment), respectively will be used in the above derivation. For the Extension Phase, the date of last dose of IMP in the Extension Phase will be used in the extension phase will be used in the above derivation and this data will be listed only.

During the course of the study, the dose may be interrupted due to safety concerns. The start date and end date of the interruption will be recorded on the Dose Interruptions page of the CRF. The number of subjects with an interruption at any point of the study will be summarized and the duration of interruption will be calculated as:

$$Duration\ of\ Interruption\ (weeks) = \frac{(End\ Date\ of\ Interruption - Start\ Date\ of\ Interruption) + 1}{7}$$

The reasons for the interruption will also be presented. If a subject has multiple interruptions during the study period, all individual interruptions will be summed to determine the overall period of treatment interruption.

The total number of doses administered will be calculated using values from the Drug Accountability CRF form and will be defined as:

$$Total\ Number\ of\ Doses\ Administered = Sum\ of\ no\ of\ doses\ dispensed - Sum\ of\ no\ of\ doses\ returned$$

With the doses dispensed at a visit calculated as the number of bottles dispensed at a visit multiplied by the amount of tablets per bottle dispensed (70 tablets per bottle). The number of doses returned at a visit will be the sum of all tablets returned per bottle. If at a visit the number of doses returned for a bottle is missing, it will be assumed that the subject used the full amount of tablets in a bottle i.e. 0 will be used as the returned amount in the calculation. The total number of doses administered will be summarized for the Main Part, Extension Phase and Overall.

The compliance to study drug will also be calculated as:

$$Compliance\ (\%) = 100 * \frac{Total\ Number\ of\ Doses\ Administered}{Expected\ number\ of\ tablets\ to\ be\ taken}$$

The expected number of tablets to be taken will be estimated based on the duration of exposure and defined as:

- (Date of last dose in the period of interest – date of first dose in the study +1) X 4 tablets/day

The following compliance categories will also be displayed: <75%, ≥75 to <90%, ≥90%, 100%. Compliance and total number of doses administered will be calculated for the SS (by treatment period defined in Section 10.7.1) and Extension Phase (where applicable), where expected number of tablets will be the expected number during each reporting phase and total number of doses administered the number administered during each phase.

Study drug accountability, study drug compliance and interruptions will be listed.

12.4 Important Protocol Deviations

Per ICON processes, protocol deviations data will be entered into the ICON system of record (PSO), in accordance with the Protocol Deviation Guidance Document. Important protocol deviations are defined in the protocol deviation guidance document. The final approved version of protocol deviation guidance will be finalized before the database lock for the primary analysis. The study team and the Sponsor will conduct on-going reviews of the deviation data from PSO, adjusting the deviation criteria as seems appropriate.

Protocol deviation data will be reviewed prior to database lock and important deviations with a potential impact on efficacy leading to the elimination of subjects from the PP will be identified. The PP must be finalized at the blinded data review meeting (or earlier), prior to database lock for the primary analysis and agreed with the Sponsor.

During the blinded data review meeting that was conducted prior to database lock, it was identified that several important deviations with impact on efficacy per section 9.6.4 were not entered into PSO. As such, programmatic PDs for deviations related to inclusion/exclusion criteria and non-compliance with study drug intake (≤80%) were derived and agreed in the blinded data review meeting and subsequently used as an additional protocol deviation source for analysis.

The number of subjects with important protocol deviations will be summarized in the FAS by category of violation and by treatment group including subset of Important protocol deviations leading to exclusion from the PP. For the primary analysis, all protocol deviations will also be listed. Deviations that occurred over a 24-weeks treatment period, over a 52-weeks treatment period and occurred during the extension phase will be flagged in the listings.

12.5 Efficacy Analyses

12.5.1 Hypothesis Testing Strategy and Multiplicity

The objective of the primary efficacy analysis will be to evaluate the efficacy of setanaxib versus placebo in the treatment of subjects with PBC and elevated liver stiffness. The aim of the efficacy analysis is to demonstrate superiority of setanaxib over placebo.

The null hypothesis is no difference in change in ALP (%) between setanaxib and placebo over 24 weeks. The alternative hypothesis is that setanaxib demonstrates a difference in change in ALP (%) over placebo.

To control the study-wise error rate, that is, the probability of rejecting the true null hypothesis for the primary endpoint at 1-sided 0.025, a Hochberg step-up procedure will be applied to adjust for 2 dose comparisons versus placebo. The hypothetical example below is used to illustrate the procedure.

- The primary analysis will be performed and the p-values for each setanaxib dose versus placebo will be assigned a rank (#1 for the smallest p-value and #2 for the largest p-value).

- Next, the Benjamini-Hochberg (BH) critical value will be calculated for each p-value, using the formula: (i/m)*Q where *i* = rank of p-value (1 or 2), m is the number of tests (this will be set to 2) and Q is the false positive risk (this will be set to 0.025, for a 1-sided test). Thus, the p-value with Rank 1 will have a BH critical value of 0.0125 and the p-value with Rank 2 will have a BH critical value of 0.025.

  Example

Comparison	Hypothetical observed P-value	Rank	BH-critical value (i/m)*Q
1. Setanaxib 1600mg vs placebo	0.020	1	0.0125
2. Setanaxib 1200mg vs placebo	0.024	2	0.025

- The test with the largest p-value that is less than its Benjamini-Hochberg critical value is comparison #2. This test and all tests with a smaller p-value will be considered significant. Therefore, in this hypothetical example, both dose comparisons will be deemed statistically significant.

For the analysis of secondary and exploratory endpoints, multiplicity will not be adjusted. “Nominal” p-values will be presented and will not be considered as part of the study-wise control of type I error. This is considered appropriate for an exploratory Phase 2b study.

12.5.2 Primary Estimand

For the primary efficacy analysis, each setanaxib treatment group (1600 mg and 1200 mg) will be compared with placebo. The primary estimand is to assess the change in ALP (%) over 24 weeks in subjects with PBC, elevated liver stiffness, and intolerance or inadequate response to UDCA.

The intercurrent events of Important protocol deviations impacting efficacy and treatment discontinuation prior to Week 24 for any reason not considered under the Composite Strategy (below) will be handled with a treatment policy strategy (the primary estimand) and the value recorded for the variable of interest is used regardless of whether or not the intercurrent event occurs (ICH E9 R1 addendum). The intercurrent events of starting additional PBC-related medications after randomization and prior to Week 24 or death prior to week 24 will be handled with a composite variable strategy, where efficacy data collected following the intercurrent event will be excluded from the primary analysis. PBC-related medication(s) will be identified through blinded medical review prior to database lock.

Table 6: Definition of Intercurrent Events

Intercurrent Event	**Strategy**	**Comment**
Important protocol deviations that impact efficacy	Treatment policy strategy	The occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest, the value for the variable of interest is used regardless of whether or not the intercurrent event occurs
Starting additional PBC-related medications after randomization and before Week 24	Composite variable strategies	Efficacy data collected after these events will be excluded from the primary analysis.
Death before Week 24		

12.5.2.1 Primary Analysis

Change in ALP at each scheduled timepoint up to Week 24 will be explored via tabulation and graphs for the FAS. Counts for the number of subjects that have a ≥15%, ≥30%, ≥40%, ≥50% and ≥70% reductions in ALP from baseline will be presented (these counts will be cumulative i.e a subject with ≥30% reduction at Week 24 will be counted in both ≥15% and ≥30% categories. Summary statistics for each schedule timepoints will be presented. These will include the geometric mean, geometric mean +/- geometric

standard deviation, median, minimum, maximum. Baseline will be summarized in two ways – including all baseline data and baseline recalculated for subjects with evaluable Week 24 results (i.e. the subset of patients who have both baseline and week 24 ALP results).

In order to summarize % change at each scheduled timepoint, the ratio of ALP to baseline will be calculate for each scheduled timepoint and a logarithmic transformation applied. The mean and mean +/- standard deviation and mean +/- standard error of the log ratios will be calculated. The results will be back-transformed to get the geometric mean ratio (+/- geometric standard deviation and +/- geometric SE).

These results will be converted to the % scale and plotted as a line graph for each treatment group. ALP values after initiation of PBC-related medication will be treated as missing values at the analysis visit following initiation of the medication.

For the primary analysis, the ratio of ALP at week 24 to baseline will be log-transformed, as will ALP data from other timepoints used in the model. This will be analyzed using a Mixed Model Repeated Measures (MMRM) analysis based on the FAS and incorporating ALP data from baseline, weeks 2,4,8,12 and 24. Baseline ALP will be included as a covariate and will be calculated as the geometric mean of pre-randomization ALP results, as defined in section 10.1 and log-transformed prior to inclusion in the analysis model. The model will also include terms for treatment group, visit, log(baseline) by visit, and visit by treatment group interaction. Subject will be included as a random effect.

the Mixed Model Repeated Measures (MMRM) model can be mathematically represented as follows:

$$Y_{ij} = \beta_0 + \beta_1 \cdot \text{ Treatment }_i + \beta_2 \cdot \text{ Visit }_j + \beta_3 \cdot \left(\log\left(\text{ Baseline ALP }_{ij}\right)\right) +$$
$$\beta_4 \cdot \left(\log\left(\text{ Baseline ALP }_{ij} \times \text{ Visit }_j\right)\right) + \beta_5\left(\text{ Treatment }_i \times \text{ Visit }_j\right) + b_i + \epsilon_{ij}$$

- Y_{ij} is the observed response (log-transformed ratio of ALP at visit j (up to Week 24) to baseline).
- β_0 is the overall intercept.
- β_1 is the effect of treatment group i.
- β_2 is the effect of visit j.
- β_3 is the effect of baseline ALP.
- β_4 is the interaction between log-transformed baseline and visit.
- β_5 is the interaction between treatment and visit.
- b_{ijk} represents the random effect for the subject k.
- ϵ_{ij} is the residual error term.

An unstructured covariance matrix will be used to model the within-subject correlation of data. The Kenward-Roger's degrees-of-freedom adjustment will be used. Restricted maximum likelihood will be used to obtain parameter estimates.

Assessment of model assumptions:

Model assumptions of the MMRM will be assessed. Where the residual errors ϵ_{ij} are assumed to be independent and identically distributed with constant variance; the random effects b_{ijk} are assumed to be normally distributed with mean zero.

1. The assumption of linearity (linear relationship between log-transformed ratio of ALP with independent variables) and homoscedasticity will be checked using residual plots.
2. The assumption for normality of residuals will be checked using QQ plots.

Assessment of model convergence:

- If the model using unstructured covariance matrix does not converge, heterogeneous Toeplitz covariance structure will be used instead of the unstructured covariance matrix. If the model still fails to converge , the Toeplitz structure will be used.
- If the above method does not lead to convergence of the model, consideration will be given to increasing the maximum number of iterations, changing the optimization technique, or adjusting the convergence criteria to promote model convergence.
- If the above optimization still fails to make the model converge, the following models (removing baseline * visit interaction term and removing the baseline term) will be used until the model converges.

$$Y_{ij} = \beta_0 + \beta_1 \cdot \text{ Treatment }_i + \beta_2 \cdot \text{ Visit }_j + \beta_3 \cdot (\log(\text{ Baseline ALP }_{ij})) + \beta_4(\text{ Treatment }_i \times \text{ Visit }_j) + b_i + \epsilon_{ij}$$

$$Y_{ij} = \beta_0 + \beta_1 \cdot \text{ Treatment }_i + \beta_2 \cdot \text{ Visit }_j + \beta_3 \cdot (\text{ Treatment }_i \times \text{ Visit }_j) + b_i + \epsilon_{ij}$$

Data presentation:

The least squares means will be estimated by visit along with the associated 95% CI. Geometric least squares mean values will be obtained by exponentiating the least squares means.

The treatment effect will be expressed as the % change in ALP for each setanaxib dose versus placebo, derived from the geometric least squares mean ratios estimated at Week 24 (and scheduled timepoints up to Week 24) interacted with each treatment arm. The 95% CI and 1-sided p-value will be presented.

To aid interpretation, corresponding estimated absolute change from baseline for each arm and the estimate absolute change versus placebo and 95% CIs will be estimated from the ratios.

12.5.2.2 Sensitivity Analyses

For continuous endpoints to be analyzed using MMRM, no explicit imputation of missing data is needed as the MMRM analysis will be performed on observed cases and implicitly imputes missing data. =

Sensitivity analyses using different ways of handling missing ALP data will be performed for the primary endpoint to assess the impact of missing data and the intercurrent strategy on the primary efficacy results:

1. Sensitivity analysis #1 While it is intended that complete data will be recorded at 24 weeks, if there is a significant amount of missing data, the robustness of the primary analysis will be assessed using multiple imputation (Section 12.5.2.2.1)
2. Sensitivity analysis #2: If examination of residual plots indicates significant outliers, then ALP data at 24 weeks will additionally be analyzed using a Robust Regression approach after first multiple imputing any missing data (Section 12.5.2.2.2)

12.5.2.2.1 Sensitivity Analysis #1 - Multiple Imputation for Sensitivity Analysis of Primary Endpoint

If missing data are observed following discontinuation, the MMRM will estimate what would have happened had patients not discontinued, and the treatment effect may be biased. To assess the robustness, a multiple imputation approach will be performed where the relationship between ALP data and time from discontinuation will be modelled within each treatment group separately, and using that relationship, data missing post-discontinuation will be imputed based on data observed post-discontinuation in other patients.

The imputation model will include data from baseline and Weeks 2, 4, 8, 12, and 24 and will impute missing ALP values as follows:

- For subjects who discontinue study treatment and who do not have subsequent ALP measurements, the missing data imputation method will use subjects in the same treatment arm who discontinue study treatment and do have subsequent ALP measurements (defined as

'retrieved dropouts') allowing for how long prior to each visit the subject discontinued. Therefore, if there is some loss of efficacy post discontinuation of therapy, subjects who discontinue study treatment and have missing data will have data imputed based on such a loss of efficacy.

- For subjects who initiate a new PBC medication, all data collected post the commencement of PBC medication will be treated as missing data in the model.
- In the unlikely event that subjects die before a visit, they will have the largest ALP value, observed across treatment arms, imputed at that visit.

The variables to be used in the analysis are described below together with sample SAS® code [Note the actual variable names may vary in the analysis]:

- TRT = treatment group
- VAL1-VAL5 = log-ALP value at each post-baseline visit for Weeks 2,4,8,12 and 24, respectively
- WDAW1-WDAW5 = Time from treatment discontinuation at Weeks 2,4,8,12 and 24, respectively. Covariates will be constructed for each scheduled visit up to Week 24 and depending on the time point being imputed, the corresponding covariate will be used in the imputation model. For example, if Week 24 is being imputed the Week 24 covariate below will be used to determine time from treatment discontinuation in the model. Each WDAW covariate takes a value of 0 if a subject is still receiving therapy at that visit and will be fitted as class variables. The value of the covariates will be derived as follows (the visit windows defined in section 10.3 will be applied):
    - WDAW1 (Week 2 Visit): If the subject discontinued treatment prior to Week 2 Visit set to 1; otherwise set to 0
    - WDAW2 (Week 4 Visit): If the subject discontinued treatment prior to Week 2 Visit set to 2; if the subject discontinued IMP on or after Week 2 Visit but prior to Week 4 Visit set to 1; otherwise set to 0
    - WDAW3 (Week 8 Visit): If the subject discontinued treatment prior to Week 2 Visit set to 3; if the subject discontinued IMP on or after Week 2 Visit but prior to Week 4 Visit set to 2; if the subject discontinued treatment on or after Week 4 Visit but prior to Week 8 Visit set to 1; otherwise set to 0
    - WDAW4 (Week 12 Visit): If the subject discontinued treatment prior to Week 2 Visit set to 4; if the subject discontinued treatment on or after Week 2 Visit but prior to Week 4 Visit set to 3; if the subject discontinued treatment on or after Week 4 Visit but prior to Week 8 Visit set to 2; if the subject discontinued treatment on or after Week 8 Visit but prior to Week 12 Visit set to 1; otherwise set to 0
    - WDAW5 (Week 24 Visit): If the subject discontinued treatment prior to Week 2 Visit set to 5; if the subject discontinued treatment on or after Week 2 Visit but prior to Week 4 Visit set to 4; if the subject discontinued treatment on or after Week 4 Visit but prior to Week 8 Visit set to 3; if the subject discontinued treatment on or after Week 8 Visit but prior to Week 12 Visit set to 2; if the subject discontinued treatment on or after Week 12 Visit but prior to Week 24 Visit set to 1; otherwise set to 0
- WD = binary covariate indicating whether the subject had discontinued from therapy prior to Week 24 Visit
- USUBJID = Subject ID
- VISIT = Visit name
- CHG = Change of log-ALP baseline ratio at each visit
- BASE = log-baseline ALP

Using the monotone datasets generated for ALP, the missing ALP data will be imputed separately for each treatment group using a REGRESSION method in conjunction with the MONOTONE statement of the SAS MI procedure. The steps are as follows:

Step 1: is to create data with a monotone data structure. For example, at this stage, data would only be imputed at Week 12 if the subject had data recorded at week 24. This allows subjects with a non-monotone pattern of missing data to be included in the analysis. The imputation model used to create monotone data pattern will include the following factors and covariates: randomized treatment group (TRT), log baseline ALP (BASE), covariate indicating if the subject discontinued treatment prior to Week 24 (WD) and the value of log-ALP at all visits up to Week 24 (VAL 1-VAL5). The WD covariate may be removed from the final model.

```
proc mi data=data nimpute=20 seed=7872228 out=partial;
        var BASE WD VAL1-VAL5;
        mcmc chain=single nbiter=200 niter=200 impute=monotone;
        by TRT;
run;
```

NBITER will be set to 200, so 200 burn-in iterations are created before sampling the first imputed dataset, and NITER also set to 200 so that each subsequent imputation is separated by 200 from the same chain. An autocorrelation plot will be produced to confirm that NITER is long enough to ensure imputations are independent and a trace plot also produced to confirm there are no systematic trends.

Step 2: Missing ALP data will be imputed separately for each treatment arm using a regression model and incorporating time from treatment discontinuation covariates as follows:

```
proc mi data=partial nimpute=1 SEED=439384 OUT=imputed;
        class WDAW1-WDAW5;
        monotone reg(VAL1=BASE WDAW1/details);
        monotone reg(VAL2=BASE WDAW2 VAL1/details);
        monotone reg(VAL3=BASE WDAW3 VAL1 VAL2 /details);
        monotone reg(VAL4=BASE WDAW4 VAL1 VAL2 VAL3 /details);
        monotone reg(VAL5=BASE WDAW5 VAL1 VAL2 VAL3 VAL4 VAL5/details);
        var BASE WDAW1-WDAW5 VAL1-VAL5;
        by TRT _imputation;
run;
```

The number of imputations will depend on the percentage of missing data to be imputed and will range from a minimum of 5 to a maximum of 50 imputed datasets. The below formula will be used to calculate the number of imputations:

$$Number\ of\ Imputations = 5 + 1.5 * (m)$$

where m is the proportion of missing data, expressed as a percentage. The smallest integer greater than or equal to the number will be used for the number of imputations.

Note, individual covariate terms for time from treatment discontinuation may be dropped from the model and/or factor levels removed if a particular pattern of missing data is not present when data are unblinded. In particular, if there are subjects present with missing data for a specific level of one of the covariates for time from treatment discontinuation, but there are no matching subjects with data observed for the same level of the same factor in the same treatment arm, the factor level will be grouped with the next highest level to enable the imputation model to be run (for example Week 2 will be grouped with Week 4). This will require running separate MI procedures if the pattern differs by treatment arm.

In the event that the above model fails, the time from treatment discontinuation covariates may be removed as follows

```
Proc mi data=partial nimpute=1 seed=439384 out=impute minmaxiter=100000;
monotone regression ;
var BASE VAL1-5;
by  TRT imputation_;
run;
```

Step 3: the MMRM model is applied to the transposed, imputed data. Any subject who dies prior to a visit will have worst case data imputed only at this stage prior to analyzing the data. Baseline ALP will be included as a covariate and will be calculated as the geometric mean of pre-randomization ALP results, as defined in section 10.1 and log-transformed prior to inclusion in the analysis model. The model will also include terms for treatment group, visit, log(baseline) by visit, and visit by treatment group interaction. Subject will be included as a random effect.

An unstructured covariance matrix will be used to model the within-subject correlation of data. The Kenward-Roger's degrees-of-freedom adjustment will be used. Restricted maximum likelihood will be used to obtain parameter estimates. Model assumptions of the MMRM will be assessed using residual plots (such as q-q plots, histograms, box plots, and scatter plots). If the model above does not converge, heterogeneous Toeplitz covariance structure will be used instead of the unstructured covariance matrix. If convergence still does not meet, the Toeplitz structure will be used.

```
proc mixed data=imputed method=reml;
        class TRT USUBJID VISIT;
        model CHG = TRT VISIT BASE BASE*VISIT VISIT*TRT /ddfm=kr;
        repeated VISIT/ sub = USUBJID type = un;
        lsmeans TRT*VISIT /cl pdiff;
        by _imputation_;
        ods output LSMeans=mixlsmean Diffs=mixdiff;
run;
```

Step 4: In the pooling phase, estimated treatment effects and associated standard errors from each imputation will be combined using Rubin's rules (Rubin, 1987) to provide an overall treatment effect, associated 95% confidence interval (CI), and 1-sided p-value.

```
proc mianalyze data=mixlsmean;
        by VISIT;
        modeleffects estimate;
        stderr stderr;
        ods output ParameterEstimates=lsmeans;
run;
```

This analysis will only be performed if there is sufficient data from 'retrieved dropouts' to ensure a reliable model with imputation.

If there are insufficient data after grouping, the imputation model may be updated to reduce the number of covariates or apply a simplified imputation method.

12.5.2.2.2 Sensitivity Analysis #2 - Robust regression for Sensitivity Analysis of Primary Endpoint

If examination of residual plots indicates clear non-normality, then ALP data at Week 24 will additionally be analyzed using a Robust Regression approach after first multiply imputing any missing data in the same manner as described in 12.5.2.2.1 Step 1. ALP values will be log-transformed prior to analysis. Results will be presented in terms of the ratio of geometric least squares mean values and their associated 95% CI.

This will be achieved by exponentiating the treatment effect and 95% CI for the mean difference in log-transformed values obtained from the robust regression model for the corresponding timepoint.

Step 1: create data with a monotone data structure having imputed 20 datasets separately within each treatment arm. The number of burn-in iterations will be set to 200, and observations will be sampled every 200 iterations within the same chain for each imputed dataset. An autocorrelation plot will be produced to confirm that NITER is long enough to ensure imputations are independent and a trace plot also produced to confirm there are no systematic trends.

```
proc mi data=data nimpute=20 seed=2019301 out=mono;
        var BASE VAL1-VAL6;
        mcmc chain=single nbiter=200 niter=200 impute=monotone;
        by TRT;
run;
```

Step 2: data will be multiply imputed using a regression method sequentially imputing data across successive visits separately by treatment arm from each dataset imputed in the first step. This analysis assumes data are missing at random conditional on previous ALP recordings, which is the same assumption used in a MMRM model and therefore would be expected to give equivalent results (O'Kelly & Ratitch).

```
proc mi data=mono nimpute=1 SEED=2019301 OUT=imputed;
        class TRT;
        var BASE VAL1-VAL5;
        monotone regression;
        by TRT _imputation_;
run;
```

log-transformed data will be analyzed using Robust Regression with independent variables of treatment and log transformed baseline ALP. M-estimation will be used with Huber weights and a cut-off value of 2 with the median method used to estimate the scale parameter. This approach means that standardized residuals with an absolute value of ≤2, corresponding to the central 95% of the data if normally distributed, have equal weight and outlying data are weighted according to a pre-specified function, $Wc\ (x)$, which gives lower weights to the most outlying data where $W_c\ (x) = min\ \{1,\ c\ /\ |x|\ \}$ where x is the standardized residual, corresponding to least square standardized residuals in the first iteration. Given that dependent variables are categorical and inclusion criteria for the only continuous covariate, log-transformed baseline ALP, prevents this variable from having outlying values, M-estimation will be appropriate.

```
proc robustreg data=data method=m (wf=huber(c=2) scale=med) outest=est1 covout;
        by _imputation_;
        class TRT;
        model CHG = BASE TRT;
run
```

In the pooling phase, estimated treatment effects and associated standard errors from each imputation will be combined using Rubin's rules to provide an overall treatment effect, associated 95% CI, and 1-sided p-value.

```
proc mianalyze data=est1;
        modeleffects estimate;
        stderr stderr;
        ods output ParameterEstimates=est2;
run;
```

12.5.2.3 Supplementary Analyses

If there are important protocol deviations that impact efficacy, the primary analysis (see Section 12.5.2.1) will be repeated using the PPS. Subjects with important protocol deviations could falsely draw the estimated treatment difference closer. Exclusion of these subjects in this sensitivity analysis will help assess the magnitude, if any, of this effect.

12.5.2.4 Summary of Analyses for the Primary Endpoint

The overview of analyses for the primary endpoint are summarized in the Table below.

Table 7: Summary of Analyses for the Primary Endpoint

Analysis	Population	Method
Summary of reductions in ALP from baseline at each scheduled time point up to Week 24	FAS	• Cumulative counts for the number of subjects with specific percentage reductions (≥15%, ≥30%, ≥40%, ≥50%, and ≥70%) from baseline • Bar chart visualization at each time point by treatment
Summary of ALP at each scheduled time point up to Week 24	FAS	• Summary Statistics (including geometric mean, geometric standard deviation, median, minimum, and maximum, and geometric mean ratio of ALP to baseline)
	FAS	• Individual line plots of ALP versus time for each subject by treatment • Line plot for geometric mean ALP vs time by treatment • Line plot for geometric mean ratio ALP (in % reduction) vs time by treatment
Primary Analysis - % change in ALP for each setanaxib dose versus placebo, derived from geometric least squares mean ratios estimated	FAS/PPS	Mixed Model Repeated Measures (MMRM)
		Forest plot of geometric mean to compare FAS, PPS, Sensitivity Analyses #1 and #2
Sensitivity Analysis #1 – Deal with missing data	FAS	Multiple Imputation (MAR) + Mixed Model Repeated Measures (MMRM) to provide an overall treatment effect, associated 95% confidence interval (CI), and 1-sided p-value
Sensitivity Analysis #2 – Assess impact of outliers	FAS	Multiple Imputation (MAR) + Robust Regression

12.5.3 Secondary Estimands

12.5.3.1 Secondary Analyses

The secondary estimands are:

- To assess the change in fatigue (as assessed by the **PROMIS-short form-Fatigue 7b Daily** and by the **PBC40 fatigue domai**n) over 24 weeks in subjects with PBC, elevated liver stiffness, and intolerance or inadequate response to UDCA. The intercurrent events are as defined for the primary estimand (see Table *5*).
- To assess the change in liver stiffness over 24 weeks in subjects with PBC, elevated liver stiffness, and intolerance or inadequate response to UDCA. The intercurrent events are as defined for the primary estimand (see Table *5*).
- To assess the change in ALP, fatigue and liver stiffness in subjects with PBC, elevated liver stiffness, and intolerance or inadequate response to UDCA, where both setanaxib doses are

combined into one setanaxib treatment group. The intercurrent events are as defined for the primary estimand (see Table *5*).

The primary analysis methodology described in Section 12.5.2.1 will be repeated for the secondary estimands, with the exception that the outcome variable for analyses of fatigue will be absolute change from baseline (*calculated as postbaseline minus baseline values*) with no log transformation and the baseline covariate in the model will not be log transformed. As described below, Y_{ij} is the observed response (absolute change in fatigue at visit j (up to Week 24) to baseline).

$$Y_{ij} = \beta_0 + \beta_1 \cdot \text{ Treatment }_i + \beta_2 \cdot \text{ Visit }_j + \beta_3 \cdot \left(\text{ Baseline fatigue }_{ij} \right) +$$
$$\beta_4 \cdot \left(\text{ Baseline fatigue }_{ij} \times \text{ Visit }_j \right) + \beta_5 \left(\text{ Treatment }_i \times \text{ Visit }_j \right) + b_i + \epsilon_{ij}$$

For the analysis of fatigue, fatigue data captured at Baseline, Weeks 4, 12, and 24 will be included in the model. Subjects enrolled under protocol version 1.0 and reconsented to version 2.0 will have no baseline value for the PROMIS-short form-Fatigue 7b as this questionnaire was introduced in version 2.0 at the request of the FDA. A sensitivity analysis will be performed with imputation of missing data.

For the analysis of change in liver stiffness at Week 24 compared to baseline, as assessed using FibroScan®, liver stiffness data captured at Baseline, Weeks 12 and 24 will be included in the model. The log transformation of liver stiffness will be performed as the same as primary analysis, the MMRM model is mathematically represented as follows, where Y_{ij} is the observed response (log-transformed ratio of liver stiffness at visit j (up to Week 24) to baseline).

$$Y_{ij} = \beta_0 + \beta_1 \cdot \text{ Treatment }_i + \beta_2 \cdot \text{ Visit }_j + \beta_3 \cdot \left(\log \left(\text{ Baseline liver stiffness}_{ij} \right)\right) +$$
$$\beta_4 \cdot \left(\log \left(\text{ Baseline liver stiffness}_{ij} \times \text{ Visit }_j \right) + \beta_5 \left(\text{ Treatment }_i \times \text{ Visit }_j \right) + b_i + \epsilon_{ij}\right.$$

For the analysis of ALP, fatigue and liver stiffness for the combined setanaxib doses, the MMRM analyses will be repeated with 2 treatment groups rather than three.

12.5.3.2 Sensitivity Analysis

The sensitivity analysis with multiple imputation defined in 12.5.2.2.1 will be performed for the secondary estimands.

12.5.3.3 Supplementary Analyses

For the combined setanaxib treatment group versus placebo, where there is sufficient data (at least 10 patients at each level of the subgroup within each treatment group), additional *exploratory* analyses examining the consistency of the treatment difference among subgroups using the FAS will be performed. The primary analysis will be repeated at each level of the subgroup. A tabulation at Week 24 will be presented for each of the following subgroups:

- Age group: ⩾18 and <65 years, ⩾ 65 years
- Sex: Female vs Male
- Actual Randomization Strata: Screening serum ALP < or ≥3.0×ULN
- UDCA use ongoing at Randomization: Yes vs No
- UDCA use prior to or at Randomization: Yes vs No
- OCA use ongoing at Randomization: Yes vs No
- OCA use prior to or at Randomization: Yes vs No
- Fibrate use (bezafibrate or fenofibrate) ongoing at Randomization: Yes vs No

12.5.3.4 Summary of Secondary Analyses

The overview of primary analyses are summarized in Table below.

Table 8: Summary of Secondary Analyses

Endpoints	Population	Analysis
Change in fatigue assessed by the PROMIS-short form-Fatigue 7b Daily (*3 treatment groups*)	FAS	• Summary Statistics of change (and % change) in the PROMIS short form fatigue 7b compared to baseline by time points • Line plot for mean of PROMIS short form fatigue 7b vs time point for by treatment
	FAS	• MMRM Model of absolute change in the PROMIS short form fatigue 7b compared to baseline
	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in fatigue assessed by the PBC40 fatigue domain (*3 treatment groups*)	FAS	• Summary Statistics of change (and % change) in the PBC40 fatigue domain compared to baseline • Line plot for mean of PBC40 fatigue domain vs time point by treatment
	FAS	• MMRM Model of absolute change in the PBC40 fatigue domain compared to baseline
	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in liver stiffness over 24 weeks (*3 treatment groups*)	FAS	• Summary Statistics of change (geometric ratio) in the liver stiffness compared to baseline • Line plot for geometric mean vs time point by treatment
	FAS	• MMRM Model of change in the liver stiffness compared to baseline
	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in fatigue assessed by the PROMIS-short form-Fatigue 7b Daily (*combined setanaxib group*)	FAS	• Summary Statistics of change (and % change) in the PROMIS short form fatigue 7b compared to baseline
	FAS	• MMRM Model of absolute change in the PROMIS short form fatigue 7b compared to baseline
	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in fatigue assessed by the PBC40 fatigue domain (*combined setanaxib group*)	FAS	• Summary Statistics of change (and % change) in the PBC40 fatigue domain compared to baseline
	FAS	• MMRM Model of absolute change in the PBC40 fatigue domain compared to baseline
	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in liver stiffness over 24 weeks (*combined setanaxib group*)	FAS	• Summary statistics for liver stiffness at each visit (including geometric mean, geometric standard deviation, median, minimum, and maximum) • Summary of geometric ratio of liver stiffness compared to baseline by time point
	FAS	• MMRM Model of change in the liver stiffness compared to baseline
	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in ALP up to Week 24 (*combined setanaxib group*)	FAS	• Summary Statistics (including geometric mean, geometric standard deviation, median, minimum, and maximum) • Summary of geometric ratio of ALP compared to baseline by time point
	FAS	• MMRM Model of change in the ALP compared to baseline

	FAS	• Sensitivity Analysis using MCMC monotone imputation
Change in ALP up to Week 24 (*combined setanaxib group*) by Subgroup • Age • Sex • Actual Randomization Strata • UDCA use ongoing at Randomization • UDCA use prior to or at Randomization • OCA use ongoing at Randomization • OCA use prior to or at Randomization • Fibrate use (bezafibrate or fenofibrate) ongoing at Randomization	FAS	• MMRM Model of change in the ALP compared to baseline (*performed at least 10 patients at each level of the subgroup within each treatment group*)

12.5.4 Other Secondary Endpoints Analyses

For the secondary endpoints listed below, summary statistics for each scheduled timepoint will be presented. These will include the arithmetic mean, standard deviation, geometric mean, geometric mean +/- geometric standard deviation, median, minimum, maximum. For the % change, if the data are lognormally distributed, the ratio of post-baseline result to baseline will be calculate for each scheduled timepoint and a logarithmic transformation applied.

The mean and mean +/- standard deviation and mean +/- standard error of the log ratios will be calculated. The results will be back-transformed to get the geometric mean ratio (+/- geometric standard deviation and +/- geometric SE for graphical purposes). These results will be converted to the % scale and plotted as a line graph for each treatment group. Values after initiation of PBC-related medication will be treated as missing values at the analysis visit following initiation of the medication. Summary tables will include scheduled visit out to Week 52 + EoT (noting that the number of subjects will be small beyond week 24), but graphs will display data for the 24-week treatment period only.

- Change in PGIS fatigue at Week 24 compared to baseline
- PGIC fatigue by time point
- Change in WI-NRS compared to baseline
- Change in PBC-40 itch compared to baseline
- Change in PGIS pruritus compared to baseline
- PGIC pruritus by time point

12.5.4.1 Markers of Cholestasis at Week 24

To summarize the changes in markers of cholestasis, the number and percentage of subjects at each scheduled timepoint meeting the following criteria for FAS will be presented:

- ALP reduction to <1.67×ULN and total bilirubin ≤1xULN and a ≥15% or ≥30% or ≥40% or ≥70% ALP reduction from Baseline, respectively
- ALP reduction to <1.5×ULN and total bilirubin ≤1xULN and with a ≥40% ALP reduction from Baseline
- ALP reduction to <1xULN and total bilirubin reduction to ≤1xULN
- Total bilirubin reduction to <0.6×ULN

Descriptive proportions will be presented for all scheduled time points up to Week 24. Subjects who start a PBC-related medication will be considered a non-responder at the time point after initiation of the PBC medication. Subjects who experience and adverse clinical outcome prior to Week 24 will also be considered non-responders.

12.5.5 Exploratory Endpoint Analysis

Exploratory efficacy endpoint analysis will be non-inferential in nature. The following endpoints will have summary statistics produced All analysis will be performed based on the observed case i.e. only subjects with data at the visit will be summarized in the analysis, which will be based on the FAS.

12.5.5.1 Change in Liver and Spleen Stiffness

Any data collected for liver and spleen stiffness measured using MRE (Protocol versions 1.0 to 4.0 only) will be listed, including change from baseline values. Any results collected during the extension phase will be flagged.

12.5.5.2 Change in Patient-Reported Outcomes

The change from baseline over the 24 week treatment period for the following patient-reported outcomes will be summarized:

- PBC 40 domains: Social Function, Cognitive Function, Emotional function, Symptoms (refer to Section 10.12)
- Overall health-related quality of life as assessed by the PROMIS-29 questionnaire (refer to Section 10.11).
- The MMRM analysis will be conducted for PROMIS-29 fatigue domain in the same manner as the change in fatigue (refer to Section 12.5.3.1). No sensitivity analysis will be performed.
- EQ-5D utility index (refer to Section 10.16)

Only summary statistics for the absolute and change from baseline scores will be presented.

12.5.6 Additional exploratory analyses

Additional exploratory analyses may be performed by the Sponsor to aid data interpretation:

- Scatter plots to show the correlation between QoL scores for fatigue at baseline versus change from baseline at week 24.
- Scatter plots to show the correlation between QoL scores for fatigue domains and AEs of fatigue.
- Scatter plots and summaries of change over time in fatigue and other QoL domains by severity of fatigue at baseline for PBC-40 (where mild <=28 and moderate/severe >=29)
- Imputation of (baseline) missing PROMIS short form fatigue 7b using the modelled relationship with the baseline PBC40 data
- Graphical presentation of changes in ALP, fatigue and liver stiffness where results from the PH2b study and the prior Ph2 study are combined in one plot.

12.6 Safety Analyses

12.6.1 Adverse Events

All AEs will be coded using MedDRA Version 24.0 or later. All AEs (including non-treatment-emergent and post-treatment events) recorded on the CRF will be listed. AEs that are not treatment emergent will be flagged for both treatment emergent definitions described in Section 10.7.1 and AEs that are assigned to the extension phase will also be flagged. Only TEAEs (see Section 10.7.2) will be included in AE summary tabulations unless otherwise specified. Summary tables of TEAEs will be presented by treatment group and

total for the SS and repeated for both treatment emergent definitions. The National Cancer Institute common terminology criteria for Adverse Events (CTCAE) toxicity grading criteria version 4.03 will be used for toxicity grading of AEs.

All AE summaries will include the number and percentage of subjects in each category and the number of events. For the calculation of the incidence of events, subjects will only be counted once within the events. For summaries by SOC and PT, subjects will be counted once per unique PT and SOC. For summaries by SOC and PT tables should be sorted in descending order of SOC and then PT by subject counts in the total group.

An overall summary of TEAEs will be presented with the following categories:

- Any TEAEs
    - TEAEs related to IMP
- Any severe TEAEs (CTCAE Grade 3 or greater)
    - Any severe TEAEs (CTCAE Grade 3 or greater) related to IMP
- Any TESAEs
    - Any TESAES related to IMP
- TEAEs with outcome of death (events with outcome of Fatal on the AE CRF page)
    - TEAEs related to IMP with outcome of death
- TEAEs leading to discontinuation of study treatment
    - TEAEs related to IMP leading to discontinuation of study treatment
- TEAEs leading to modification of study treatment
    - TEAEs related to IMP leading to modification of study treatment
- TEAEs leading to study discontinuation
    - TEAEs related to IMP and leading to study discontinuation
- TEAEs of special interest
    - TEAEs of special interest related to IMP
- TEAEs requiring Adjudication
- TEAE related to Hepatic Encephalopathy (as indicated by the Investigator on the AE CRF page)

The following summaries will be presented by SOC and PT :

- TEAEs
- TEAEs by Maximum CTCAE Grade
- TEAEs related to IMP
- TEAEs related to IMP by Maximum CTCAE Grade
- TEAEs of special interest
- TEAEs of special interest related to IMP
- TEAEs leading to discontinuation of study treatment
- TEAEs leading to study discontinuation
- TESAEs

- TESAEs related to IMP
- Severe TEAEs of CTCAE Grade 3 or higher
- Non-Serious TEAEs
- TEAEs occurring in 2 or more of subjects (based on PT) in any treatment arm and summarized by PT only in descending order

Subjects with multiple events within a particular SOC or PT will be counted under the category of their most drug-related event within that SOC or PT (first order by drug-related).

Subjects with multiple events within a particular SOC or PT will be counted under the category of their most severe event within that SOC or PT (subsequent order by severity).

A listing of all AEs will be produced with treatment emergent flags for each treatment emergent definition (Sections 10.7.1 and 10.7.2). Separate listings of TEAEs leading to death, leading to IMP discontinuation, CTCAE grade 3 or higher TEAEs, AESIs and SAEs will be provided.

12.6.2 Adjudicated Events

The following adverse clinical outcomes will be adjudicated during the study:

- All-cause mortality/Death
- Variceal bleed
- Portal hypertension bleed
- Hepatic encephalopathy (West Haven criteria)
- Spontaneous bacterial peritonitis
- Uncontrolled ascites requiring treatment

Additionally, events that require IMP interruption (DILI, anemia and hypothyroidism) will also be adjudicated.

The proportion of subjects with adjudicated events in the above categories will be summarized in the SS. All subjects with adjudicated events will be listed.

An additional summary will be produced summarizing the proportion of subjects experience the following (not necessarily adjudicated) Safety events.

- progression to cirrhosis
- death (all-cause)
- liver transplant
- MELD ≥15 if the baseline MELD score was <12
- new onset ascites requiring treatment (diuretics or paracentesis),
- development of large varices if no varices at baseline,
- hospitalization due to a decompensatory event (for example, variceal bleeding, hepatic encephalopathy, spontaneous bacterial peritonitis)

If no such events are reported, this table will not be included.

12.6.3 Laboratory Data

Laboratory data will be reported in Système International (SI) units.

For clinical biochemistry, hematology and liver function tests (Table 8), shift tables from baseline to worst CTCAE grade on treatment will be produced for the SS, with separate summaries for each treatment period. Only parameters which have toxicity grading will be summarized. WBC and reticulocyte

counts will be expressed in absolute values. Differential count will be expressed as both absolute count and percentage of WBCs. Only central lab data will be included in the table summaries. All lab data collected locally and centrally will be listed.

Table 8: Laboratory Tests

Laboratory Tests	Parameters
Hematology	Hematocrit; hemoglobin; absolute and relative reticulocyte counts, red blood cell (RBC) count; white blood cell (WBC) count; differential WBC count; platelet count; absolute neutrophil count; mean cell volume, INR
Biochemistry	Glucose; total protein; creatinine; urea; total cholesterol; triglycerides; sodium; potassium; chloride
Liver Function	Serum fibrinogen, albumin, ALP, ALT, AST, total and conjugated bilirubin, GGT, hsCRP, IgM

To further examine the liver function tests collected during the study, suspected drug-induced liver injury (sDILI) summaries will be produced. Proportion of subjects meeting any of the following criteria for any post-baseline assessments will be presented for the SS, for each treatment period:

- AST or ALT ≥3xULN
- AST or ALT >2xBaseline
- Total bilirubin >2xULN
- INR>1.5

Any subjects with any of the above criteria for sDILI or cholestasis will be flagged in the listings. Subjects at risk of a drug induced liver injury according to Hy's law will also be flagged. Hepatic function abnormality defined by an increase in AST and/or ALT to ≥ 3xULN concurrent with an increase in total bilirubin to ≥ 2xULN but without increase in ALP (i.e., ALP < 2xULN) meets the criteria for Hy's law and raises the concern for DILI.

The MELD score will be calculated by the central laboratory based on the results from the liver biochemistry (total bilirubin), biochemistry (creatinine), and hematology (INR) blood samples. Subjects with a MELD score of ≥ 15 will be flagged in the listings.

Data from the Extension Phase will be restricted to the period while a subject was on active and change from baseline comparisons using the extension baseline.

Table 9: Other Laboratory Tests

Laboratory Tests	Parameters
Markers of Liver Fibrosis	ELF score components for calculation of ELF score (TIMP-1, PIIINP, and hyaluronic acid) and the other PBC-related biomarkers (PRO-C3 and C3M; PRO-C3/CRM ratio)
Markers of bile acid metabolism	Total bile acids, C4, FGF19, FGF21
Autoantibodies	AMA titer, PBC-specific antibodies (anti-GP210, anti-SP100, antibodies against major M2 components [PDC-e2, 2-oxo-glutaric acid dehydrogenase complex])

Laboratory Tests	Parameters
Viral Serology	HIV antibodies (1 and 2) and, if positive for HIV antibodies, HIV RNA; Hepatitis B surface antigen and hepatitis C virus antibodies and, if positive for HCV antibody, HCV RNA
Pregnancy test	Serum pregnancy test; urine Pregnancy test
Thyroid function	TSH and free T4
Drug Screen (Urine)	Urine drug screen
Urinalysis	pH; protein; glucose; ketones; bilirubin; blood; microscopic examination of the sediment

Refer to Table 9*8*, for a list of other laboratory parameters to be collected on this study and the laboratory test category each parameter will be summarized under. Autoantibodies, viral serology, pregnancy test, and drug screen parameters will be listed only. Markers of liver fibrosis, markers of bile acid metabolism, urinalysis, thyroid function will be summarized descriptively by scheduled visit using the SS.

Standard ranges from the central laboratory will be used for the laboratory analysis. The assessment as to if a lab test is low, high, normal as applicable will be derived using the standard ranges for central lab data. Laboratory data outside study-specific reference ranges will be listed.

Abnormal laboratory values, defined as values outside the normal range, will be listed by subject. Summaries of the incidence and frequency by treatment group and scheduled visit will be presented.

12.6.4 eDISH Analysis

Evaluation of drug-induced serious hepatotoxicity (eDISH) analysis will be conducted in order to explore potential liver toxicity signals by treatment group. A eDISH plot will be produced, which will display for each subject during the treatment period the maximal total ALT/AST value (expressed as a multiple of the ULN) against the maximal Total Bilirubin value (expressed multiple of the ULN) to identify possible values above Hy's law range. The plot will be produced using the SS and treatment period of 52 weeks (Section 10.7.1) and will display the maximal on-treatment value of ALT/AST and total bilirubin.

12.6.5 Vital Signs

Observed values and change from baseline in vital signs will be summarized descriptively by scheduled visit and treatment group using the SS and treatment period of 52 weeks. Shift plots of baseline as compared to highest and lowest on treatment value by treatment group will be presented. Vital sign parameters to be summarized descriptively are pulse rate, systolic blood pressure (SBP), and diastolic blood pressure (DBP). By-subject listings of all the vital signs will also be provided. Body temperature will be listed only.

12.6.6 Physical Examinations, ECGs, and Other Observations Related to Safety

ECGs will be read centrally and ECG mean Heart Rate (HR), RR interval, PR Interval, QRS Duration, QT interval, QTcB interval and QTcF interval will be collected and summarized. In the Extension Phase, ECG may be read locally and HR, PR interval, RR interval, QRS interval, QT interval and QTcF interval will be listed. ECGs will be presented by the category normal, abnormal, unevaluable, and unknown for each assessment time point in the safety analysis, over the treatment period of 52 weeks. Shift plots of baseline as compared to highest and lowest on treatment value by treatment group will be presented for both treatment periods defined in Section 10.7.1. A listing of subjects ECG data will be provided.

QTc outliers are defined as QTc values following dosing that are greater than 450 msec or are increases from baseline greater than 30 msec. QTc outliers will be highlighted in the data listings, with subjects meeting the criteria at any time point listed separately. Outliers will be summarized using separate shift tables for each treatment period defined in Section 10.7for the following categories:

- Values ≤450 msec, >450-480 msec, >480-500 msec, >500 msec at baseline versus values ≤450 msec >450-480 msec, >480-500 msec, >500 msec post-dose
- Values ≤450 msec, >450-480 msec, >480-500 msec, >500 msec at baseline versus increase from baseline of ≤30 msec, >30-60 msec, >60-90 msec or >90msec post-dose

A summary of physical examination findings will be presented for each scheduled visits, showing the proportion of subjects with “Normal” and “Abnormal” physical examinations. The number of abnormal, clinically significant and not clinically significant assessments will be tabulated by scheduled visit using the SS and treatment period of 52 weeks. All physical examination findings will be provided in listings.

The Child-Pugh score and Child-Pugh classification of Severity of Cirrhosis as collected on the CRF page will be listed for all scheduled visits with flags to indicate in the value is treatment for each treatment periods defined in Section 10.7.1. The individual components required to calculate the Child-Pugh score and class will also be listed).

12.6.7 Liver Biopsy and Liver Histology

For qualifying subjects in the SS, the absolute and change from baseline in the Nakamuna total score and fibrosis subscore at Week 52 will be summarized. A listing for the historical liver biopsy and liver biopsy data will be provided.

12.7 Pharmacokinetics Analyses

Blood samples will be collected for measurements of plasma concentrations of setanaxib and its metabolite GKT138184 during the Main Part of the study. Samples will be collected pre-dose at Week 6, Week 24 and (under protocol versions 1.0 to 4.0) also Week 52. The PK analysis set will be used for PK analysis.

Plasma concentrations of setanaxib and GKT138184, along with blood sampling dates and actual blood sampling time relative to previous dosing time, will be listed by subject, dose group, actual treatment (can change within subject), by formulation; can change within subject] and nominal sampling time. Samples with time deviation from nominal time will be identified and listed. Plasma setanaxib and GK138184 concentrations below the quantifiable limit (BQL) will be set to ½ the lower limit of quantification (LLOQ) in the computation of mean concentration values.

Descriptive statistics (number of subjects, arithmetic mean, geometric mean, standard deviation, coefficient of variation for geometric mean, median, Q1, Q3, minimum, and maximum) will summarize the plasma concentrations of both setanaxib and GKT138184 at pre-dose Week 4, pre-dose Week 24 and, where relevant, pre-dose Week 52, by actual treatment (may change within subject). If 50% of the subjects at a given time point have values BQL then the descriptive statistics will not be presented and will instead display as BQL for the mean and minimum, with the exception of maximum all other statistics will be missing. Plasma setanaxib and GKT138184 concentrations (dose normalized as appropriate) vs study visit will be plotted as a box plot, by actual treatment (may change within subject) with corresponding sample sizes.

12.8 Data Collected Under Protocol Version 1.0

The following assessments will be collected under protocol version 1.0 and will not be included in the primary or final analysis. Listings will be produced for all subjects with data collected for these assessments:

- 5-D Itch Scale
- Pruritus-Visual Analogue Scale.

13.0 References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to mulitple testing. Journal of the Royal Statistical Society, Series B, 57:289-300.

Cella D, Riley W, Stone A, et al. Initial Adult Health Item Banks and First Wave Testing of Patient-Reported Outcomes Measurement System (PROMISTM) Network: 2005-2008. J Clin Epidemiol. 2010 Nov;63(11):1179-94.

EuroQol Research Foundation. EQ-5D-5L User Guide. 2019. Version 3.0 dated Sept 2019. Available at: https://euroqol.org/publications/user-guides/. Last Accessed on 21 Mar 2021.

HealthMeasures. PROMIS ADULT PROFILE INSTRUMENTS. A brief guide to the PROMIS® Profile instruments for adult respondents. https://staging.healthmeasures.net/images/PROMIS/manuals/PROMIS_Adult_Profile_Scoring_Manual.pdf. Published January 17, 2019. Accessed August 6, 2021.

HealthMeasures. FATIGUE. A brief guide to the PROMIS® Fatigue instruments. https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Fatigue_Scoring_Manual.pdf. Published February 28, 2019. Accessed December 15, 2021.

Herdman M, Gudex C, and Lloyd A, et al. Development and Preliminary Testing of the New Five-level Version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727-36.

Jacoby A, Rannard A, Buck D, et al. Development, Validation and Evaluation of the PBC-40, a Disease Specific Health Related Quality of Life Measure for Primary Biliary Cirrhosis. Gut. 2005 Nov;54(11):1622-9.

ICH. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials – Guidance for Industry. https://www.fda.gov/media/148473/download. Published May 2021. Accessed June 4, 2021.

Nakanuma Y, Zen Y, Harada K, et al. Application of a New Histological Staging and Grading System for Primary Biliary Cirrhosis to Liver Biopsy Specimens: Interobserver Agreement. Pathology International 2010;60:167–74.

Ratitch B, Lipkovich I, O'Kelly, M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG. 2013

Reich A, Heisig M, Phan NQ, et al. Visual Analogue Scale: Evaluation of the Instrument for the Assessment of Pruritus. Acta Derm Venereol. 2012 Sep;92(5):497-501.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: J Wiley & Sons, 1987.

14.0 Glossary of Abbreviations

Glossary of Abbreviations:	
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AMA	Antimitochondrial Antibodies
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomic Therapeutic Classification
BMI	Body Mass Index
BQL	Below Quantifiable Limit
C3M	A Peptide of Helical Collagen Type III Degradation
CI	Confidence Interval
CMH	Cochran Mantel Haenszel
CONSORT	Consolidated Standards of Reporting Trials
CRF	Case Report Form
CTCAE	Common Terminology Criteria for Adverse Events
DBP	Diastolic Blood Pressure
DILI	Drug-Induced Liver Injury
ECG	Electrocardiogram
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EAS	Extension Safety Analysis Set
ELF	Enhanced Liver Fibrosis
ENRL	All Enrolled Analysis Set
EoT	End of Treatment
EQ-5D	EuroQol 5-Dimension Utility Index
ES	Extension Analysis Set
FAS	Full Analysis Set
FGF	Fibroblast Growth Factor
GGT	Gamma Glutamyl Transpeptidase
GM	Geometric Mean
HR	Heart Rate
hsCRP	High Sensitivity C-reactive Protein
ICH	International Council for Harmonization
IDMC	Independent Data Monitoring Committee

Glossary of Abbreviations:	
IgM	Immunoglobulin M
IMP	Investigational Medicinal Product
INR	International Normalized Ratio
IXRS	Interactive Web Response System
LLOQ	Lower Limit of Quantification
LOCF	Last Observation Carried Forward
LS	Least Squares
MAR	Missing At Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End Stage Liver Disease
MMRM	Mixed Model Repeated Measures
MRE	Magnetic Resonance Elastography
N/A	Not Applicable
OCA	Obeticholic Acid
PBC	Primary Biliary Cholangitis
PIIINP	Amino Terminal Propeptide of Type III Procollagen
PGIC	Patient's Global Impression of Change
PGIS	Patient's Global Impression of Severity
PK	Pharmacokinetic(s)
PKS	PK Analysis Set
PPS	Per Protocol Analysis Set
PRO	Patient Reported Outcome
PRO-C3	Released N-terminal Propeptide of Type III Collagen
PROMIS	Patient-Reported Outcomes Measurement Information System
PSO	ICON System of Record
PT	Preferred Term
RBC	Red Blood Cell
SAP	Statistical Analysis Plan
SAE	Serious Adverse Event
SBP	Systolic Blood Pressure
SD	Standard Deviation
sDILI	Suspected drug induced liver injury
SE	Standard Error
SI	Système International
SOC	System Organ Class

Glossary of Abbreviations:	
SS	Safety Analysis Set
TEAE	Treatment-Emergent Adverse Event
TIMP-1	Tissue Inhibitor of Metalloproteinase 1
UDCA	Ursodeoxycholic Acid
ULN	Upper limit of Normal
WBC	White Blood Cell
WI-NRS	Worst Itch Numerical Rating Scale
WHODRUG	World Health Organization Drug Dictionary

Appendix

Appendix 1: Handling of Partial Adverse Event Dates

In order to assign an AE with partial dates as treatment emergent in the Main Part and Extension Phase the following imputation rules will be used.

Table 9: Handling of Partial AE dates (D = Day, M = Month, Y=Year)

Date of Interest	Missing	Condition	Imputation
AE start date	D	M and Y is the same as M and Y of first dose of IMP	Date of first dose of IMP
		Y is prior to Y of first dose of IMP	Last day of the month
		Y is the same as Y of first dose of IMP, M is prior to M of first dose of IMP	Last day of the month
		M and/or Y is after first dose of IMP	First day of the month
	D, M	Y is the same as the Y of first dose of IMP	Date of first dose of IMP
		Y is prior to Y of first dose of IMP	Use 31st of December
		Y is after Y of first dose of IMP	Use 1st of January
	M	Treat day as missing and use imputation rules for D and M as missing	See D, M imputation rules
	D, M, Y	None-date is completely missing	Date of first dose of IMP
AE End Date	D	M and Y same as date of last dose of IMP	Date of last dose of IMP
		M and Y not the same as date of last dose of IMP/date of last dose of IMP is missing	Set to the last day of the month, or the end of study/death date whichever is earliest
	D, M	Y is the same as last dose of IMP	Date of last dose of IMP
		Y is not the same as date of last dose of IMP/date of last dose of IMP is missing	Set to 31st of Dec, or the end of study/death date whichever is earliest
	M	Treat day as missing and use imputation rules for D and M as missing	See D, M imputation rules
	D, M, Y	None	No imputation, assume Ongoing

If the above imputation rules result in illogical dates adjust accordingly. If any estimated AE start date is after a complete/imputed AE end date, set the AE start date to be the first day in the AE end date month. If any estimated AE end date is prior to a complete/imputed AE start date set to be the last day of the AE start date month. Imputed values will not be displayed in listings.

Appendix 2: Handling of Partial Medications Dates

In order to assign medications as either prior and/or concomitant the following rules will be used.

Table 10: Handling of Partial Medication dates (D = Day, M = Month, Y=Year)

Date of Interest	Missing	Condition	Imputation
Medication start date	D	M and Y is the same as M and Y of first dose of IMP	Date of first dose of IMP
		M and/or Y is not the same as the first dose of IMP	First day of the month
	D, M	Y is the same as the Y of first dose of IMP	Date of first dose of IMP
		Y is not the same as first dose of IMP	Use 1st of January
	M	Treat day as missing and use imputation rules for D and M as missing	See D, M imputation rules
	D, M, Y	None-date is completely missing	Assume prior and concomitant. Impute as the last day prior to the first dose of IMP.
Medication End Date	D	M and Y same as date of last dose of IMP	Date of last dose of IMP
		M and Y not the same as date of last dose of IMP/date of last dose of IMP is missing	Set to the last day of the month, or the end of study/death date whichever is earliest
	D, M	Y is the same as last dose of IMP	Date of last dose of IMP
		Y is not the same as date of last dose of IMP/date of last dose of IMP is missing	Set to the 31st of Dec, or the end of study/death date whichever is earliest
	M	Treat day as missing and use imputation rules for D and M as missing	See D, M imputation rules
	D, M, Y	None	No imputation, assume Ongoing

If the above imputation rules result in illogical dates adjust accordingly. If any estimated medication start date is after a complete/imputed medication end date, set the medication start date to be the first day in the medication end date month. If any estimated medication end date is prior to a complete/imputed medication start date set to be the last day of the medication start date month. Imputed values will not be displayed in listings. For PBC medications collected on the concomitant medication page of the CRF with missing dates that are identified as impacting efficacy through medical review, the above rules will be used to impute partial dates and all assessments collected past the imputed partial dates will be seen as assessments collected while on rescue medication.

Appendix 3: PROMIS-Short Form-Fatigue 7b Daily Scoring Tables

Refer to HealthMeasures (2019) for further detail.

Fatigue 7b Daily – Adult v1.0		
Short Form Conversion Table		
Raw Score	T-score	SE*
7	31.0	4.9
8	36.3	3.5
9	39.4	3.0
10	41.6	2.7
11	43.5	2.5
12	45.2	2.4
13	46.7	2.4
14	48.1	2.3
15	49.5	2.3
16	50.9	2.3
17	52.2	2.3
18	53.6	2.4
19	54.9	2.4
20	56.2	2.4
21	57.6	2.4
22	58.9	2.4
23	60.3	2.4
24	61.7	2.4
25	63.0	2.4
26	64.4	2.4
27	65.7	2.4
28	67.2	2.4
29	68.6	2.4
30	70.2	2.4
31	71.8	2.5
32	73.6	2.6
33	75.6	2.9
34	78.1	3.2
35	81.4	3.6

*SE = Standard Error on T-score metric

Appendix 4: PROMIS-29 V2.1 Scoring Tables

Refer to HealthMeasures (2019) for further detail.

Adult v2.0 - Physical Function 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	22.5	4.0
5	26.6	2.8
6	28.9	2.5
7	30.5	2.4
8	31.9	2.3
9	33.2	2.3
10	34.4	2.3
11	35.6	2.3
12	36.7	2.3
13	37.9	2.3
14	39.2	2.4
15	40.5	2.4
16	41.9	2.5
17	43.5	2.6
18	45.5	2.8
19	48.3	3.3
20	57.0	6.6
*SE = Standard Error on T-score metric		

Adult v1.0 - Anxiety 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	40.3	6.1
5	48.0	3.6
6	51.2	3.1
7	53.7	2.8
8	55.8	2.7
9	57.7	2.6
10	59.5	2.6
11	61.4	2.6
12	63.4	2.6
13	65.3	2.7
14	67.3	2.7
15	69.3	2.7
16	71.2	2.7
17	73.3	2.7
18	75.4	2.7
19	77.9	2.9
20	81.6	3.7
*SE = Standard Error on T-score metric		

Adult v1.0 - Depression 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	41.0	6.2
5	49.0	3.2
6	51.8	2.7
7	53.9	2.4
8	55.7	2.3
9	57.3	2.3
10	58.9	2.3
11	60.5	2.3
12	62.2	2.3
13	63.9	2.3
14	65.7	2.3
15	67.5	2.3
16	69.4	2.3
17	71.2	2.4
18	73.3	2.4
19	75.7	2.6
20	79.4	3.6
*SE = Standard Error on T-score metric		

Adult v1.0 - Fatigue 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	33.7	4.9
5	39.7	3.1
6	43.1	2.7
7	46.0	2.6
8	48.6	2.5
9	51.0	2.5
10	53.1	2.4
11	55.1	2.4
12	57.0	2.3
13	58.8	2.3
14	60.7	2.3
15	62.7	2.4
16	64.6	2.4
17	66.7	2.4
18	69.0	2.5
19	71.6	2.7
20	75.8	3.9
*SE = Standard Error on T-score metric		

Adult v1.0 - Sleep Disturbance 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	32.0	5.2
5	37.5	4.0
6	41.1	3.7
7	43.8	3.5
8	46.2	3.5
9	48.4	3.4
10	50.5	3.4
11	52.4	3.4
12	54.3	3.4
13	56.1	3.4
14	57.9	3.3
15	59.8	3.3
16	61.7	3.3
17	63.8	3.4
18	66.0	3.4
19	68.8	3.7
20	73.3	4.6
*SE = Standard Error on T-score metric		

Adult v1.0 – Ability to Participate in Social Roles and Activities 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	27.5	4.1
5	31.8	2.5
6	34.0	2.3
7	35.7	2.2
8	37.3	2.1
9	38.8	2.2
10	40.5	2.3
11	42.3	2.3
12	44.2	2.3
13	46.2	2.3
14	48.1	2.2
15	50.0	2.2
16	51.9	2.2
17	53.7	2.3
18	55.8	2.3
19	58.3	2.7
20	64.2	5.1
*SE = Standard Error on T-score metric		

Adult v1.0 - Pain Interference 4a		
Short Form Conversion Table		
Raw Summed Score	T-score	SE*
4	41.6	6.1
5	49.6	2.5
6	52.0	2.0
7	53.9	1.9
8	55.6	1.9
9	57.1	1.9
10	58.5	1.8
11	59.9	1.8
12	61.2	1.8
13	62.5	1.8
14	63.8	1.8
15	65.2	1.8
16	66.6	1.8
17	68.0	1.8
18	69.7	1.9
19	71.6	2.1
20	75.6	3.7
*SE = Standard Error on T-score metric		