Statistical Analysis Plan (SAP)

Protocol Title:	A Double-Blind, Randomized, Placebo-Controlled, Phase 3 Study to Demonstrate Efficacy and Safety of A4250 in Children with Progressive Familial Intrahepatic Cholestasis Types 1 and 2 (PEDFIC 1)
Protocol Number:	A4250-005
IND number	130591
EudraCT number	2017-002338-21
Protocol Version, Date	Final Amendment 06, 24 June 2019
Document Version, Date:	Final v1.2, 25-Aug-2020
Prepared by:	[redacted]

SIGNATURE PAGE

Prepared by:

[redacted] Date

Reviewed by:

[redacted] Date

Reviewed and Approved by:

[redacted] Date

[redacted] Date

REVISION HISTORY

Version/Date	Modified by	Reason for Modifications
Final v1.0/23Oct2019	NA	
Final v1.1/03Aug2020	[redacted]	1. To add the handling for patients whose baseline s-BA ≤70 µmol/L and for patients whose baseline pruritus score is ≤ 1. 2. To add responder analysis based on the clinically meaningful threshold by using monthly scores. To add the result of clinically meaningful change from the blinded psychometric analysis. 3. To clarify off-treatment pruritus assessments will be included for analysis. For completers, the denominator to calculate US primary endpoint will be based on the last eDiary report day. 4. To update dictionary version. 5. To clarify the worst scenario will be used for tipping point analysis of EU and ROW primary endpoint. 6. To clarify the growth analysis will be based on derived growth data using the software or methods from the Centers for Disease Control and Prevention (CDC) website for patients with age ≥ 2 years old and from the WHO website for patients with age < 2 years old. Also, to add a MMRM analysis. 7. To clarify the counting of number of adverse events. To add the adverse events of interest and their definitions. 8. To clarify that local laboratory data will be used when central laboratory data are not available. 9. To clarify the calculation of change from baseline for AM&PM pruritus scores (daily, weekly, bi-weekly, monthly) for analysis purposes. 10. To update the analysis windows of assessments.

Final v1.2/25Aug2020		1, To remove completer analysis set and its analysis. 2. To add MMRM analysis for s-BA and ALT. 3. To clarify the handling for treatment completers in responder analyses.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

The following abbreviations will be used within this SAP.

Abbreviation or special term	Explanation
E	Adverse Event
AFP	Alfa-fetoprotein
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
APRI	AST to Platelet Ratio Index
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BSEP	Bile Salt Export Pump
CAP	Controlled Attenuation Parameter
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus Disease 2019
CRO	Contract Research Organization
CSR	Clinical Study Report
DSMB	Data Safety and Monitoring Board
eCRF	Electronic Case Report Form
EOT	End of Treatment
EU	European Region
FAS	Full Analysis Set
eDiary	Electronic Diary
GGT	Gamma Glutamyl Transferase
GIC	Global Impression of Change
GIS	Global Impression of Symptoms
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
[redacted]	[redacted]
INR	International Normalized Ratio
IWRS	Interactive Web Response System

Abbreviation or special term	Explanation
kg	Kilogram
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MELD	Model for End-Stage Liver Disease
MI	Multiple Imputation
MMRM	Mixed-Effect Model for Repeated Measures
MNAR	Missing Not at Random
ObsRO	Observer-Reported Outcome
p-C4	Plasma 7α-hydroxy-4-cholesten-3-one Concentration
PD	Protocol Deviation
PFIC	Progressive Familial Intrahepatic Cholestasis
PELD	Pediatric End-Stage Liver Disease
PedsQL	Pediatric Quality of Life Inventory
PK	Pharmacokinetics
PMM	Pattern Mixture Model
PP	Per Protocol
PRO	Patient-Reported Outcome
PT	Preferred Term
QoL	Quality of Life
RoW	Rest of World
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAF	Safety Analysis Set
SAS	Statistical Analysis System
s-BA	Serum Bile Acid
SD	Standard Deviation
SI	International System of Unit
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TFLs	Tables, Figures and Listings
ULN	Upper Limit of Normal
US	United States

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide detailed descriptions of the statistical methods, data derivations, and data displays for study protocol A4250-005 “A Double-Blind, Randomized, Placebo-Controlled, Phase 3 Study to Demonstrate Efficacy and Safety of A4250 in Children with Progressive Familial Intrahepatic Cholestasis Types 1 and 2 (PEDFIC 1)”. The table of contents and templates for the Tables, Figures and Listings (TFLs) will be produced in a separate document.

Any deviations from this SAP will be described and justified in the Clinical Study Report (CSR). The preparation of this SAP has been based on International Conference on Harmonisation (ICH) E3 and E9 guidelines [1, 2]. All data analyses and generation of TFLs will be performed using SAS® version 9.4 or higher. The SAP will be finalized and signed off prior to locking the database and unblinding.

2 STUDY OBJECTIVES

2.1 Primary objectives

To demonstrate the efficacy of repeated daily doses of 40 µg/kg/day and 120 µg/kg/day odevixibat (A4250) in children with progressive familial intrahepatic cholestasis (PFIC) Types 1 and 2, as determined by the following:

- The proportion of patients experiencing at least a 70% reduction in fasting serum bile acid (s-BA) concentration from Baseline to end of treatment or reaching a level ≤70 µmol/L after 24 weeks of treatment

- The proportion of positive pruritus assessments at the subject level over the 24-week treatment period. A positive pruritus assessment is defined as a scratching score of ≤1 or at least a one-point drop from baseline on the Albireo Observer-Reported Outcome (ObsRO) instrument

2.2 Secondary objectives

- To evaluate the effect of odevixibat on serum alanine aminotransferase (ALT) concentration
- To evaluate the effect of odevixibat on growth
- To evaluate the effect of odevixibat on sleep disturbance
- To evaluate the effect of odevixibat on the need for surgical treatment (biliary diversion or liver transplantation)
- To assess the safety and tolerability of repeated daily doses of odevixibat for 24 weeks

3 STUDY DESIGN

3.1 General study design

This is a double-blind, randomized, placebo-controlled, Phase 3 study to investigate the efficacy and safety of 40 µg/kg/day and 120 µg/kg/day of odevixibat compared to placebo in children with PFIC Type 1 and 2.

The study includes a 35- to 56-day Screening Period followed by a 24-week Treatment Period and a 4-week Follow-Up Period. After completion of the Screening Period, eligible patients (20 per treatment group) will be randomized on Day 0 (Visit 3) in a 1:1:1 fashion to receive 40 µg/kg/day or 120 µg/kg/day of odevixibat, or a matching placebo. Study data will be periodically reviewed by an unblinded Data Safety and Monitoring Board (DSMB).
A total of approximately 60 patients with a clinical diagnosis of PFIC Type 1 or 2 that have been genetically confirmed will be randomized. The randomization will be done in blocks and stratified according to PFIC type (Type 1 and 2) and age category (6 months to 5 years, 6 to 12 years, and 13 to ≤18 years) to ensure approximate balance between dose schemes (1:1:1).

The study will target 15% of enrolled patients to be PFIC-1 patients. If enrollment of all PFIC-2 patients is completed and the 15% enrollment of PFIC-1 patients has not yet been achieved, then enrollment of PFIC-2 patients will continue to reach the total study target enrollment. An attempt will be made to randomize a minimum of 10 patients from each age category (6 months to 5 years, 6 to 12 years, and 13 to ≤18 years). A maximum of 5 patients that were treated in Study A4250-003 will be allowed to participate in this study.

Patients who complete the Treatment Period and Visit 9 (Day 168, Week 24) in this study will be invited to participate in a 72-week open-label extension study (A4250-008, EudraCT Number 2017-002325-38, subject to a separate study approval) in which all patients will receive odevixibat 120 µg/kg/day. Prior to protocol amendment 6, patients who had completed at least 12 weeks of the Treatment Period and had either no improvement or intolerable symptoms were eligible to complete the End of Treatment (EOT) Visit and were offered the opportunity to enter Study A4250-008.

There will be up to 10 clinic visits during the study and one scheduled telephone contact, as follows:

- Visit 1: Screening Period start. Patient is considered enrolled after signing the informed consent form (ICF) (Days -56 to -35)
- Visit 2: Screening Visit 2 (Days -28 to -7)
- Visit 3: Randomization Visit (Day 0)
- Telephone Contact 1: (Day 14 - Week 2)
- Visit 4: Clinic Visit (Day 28 - Week 4)
- Visit 5: Clinic Visit (Day 56 - Week 8)
- Visit 6: Clinic Visit (Day 84 - Week 12)
- Visit 7: Clinic Visit (Day 126 - week 18)
- Visit 8: Clinic Visit (Day 154 - Week 22)

- Visit 9: Clinic Visit, End of Treatment (Day 168 - Week 24)
- Visit 10: Follow-Up Visit, 28 days after last dose of study drug (Day 196). Patients continuing in Study A4250-008 will not have this follow-up visit.

Additional clinical visits may be required for patients who need direct site assistance, e.g., due to adverse event (AE) monitoring, to fulfill screening requirements, and/or for safety maintenance.

Caregivers/patients will be instructed on the daily use of the eDiary. The eDiary will include patient-reported outcome (PRO) for patients ≥8 years of age and observer-reported outcome (ObsRO) instruments for evaluation of itching (PRO), scratching (ObsRO), and sleep disturbance (PRO and ObsRO) during the Screening Period (study Days -56 to 0), throughout the 24-week Treatment Period (study Days 0 to 168), and Follow-up (Day 196). Additionally, caregivers will be requested to use the eDiary to report the time each dose of study drug is administered during the Treatment Period. All entries made in the ediary are transmitted overnight to the database which can be accessed via Trial Manager.

The study nurse will monitor eDiary compliance by routine review of the Trial Manager website. If both diary entries on a day are missing, the study nurse will call the patient/caregiver to remind them to complete all scheduled entries. Any noncompliance will be documented and explained in the source documents.

3.2 Randomization and blinding

A double-blind design is employed so that both the investigators and the patients will be unaware of the treatment assignment during the study. The sponsor and contract research organization (CRO) personnel (except CRO statisticians generating the list of randomization codes) involved in this study will also not be aware of the treatment assignment during the study.

After written informed consent/assent is obtained from an eligible patient/legal guardian, an 8-digit patient identification number will be assigned. The first 2 digits will denote the country, followed by a 3-digit site number and a 3-digit patient sequence number. This number will be created and allocated by the Interactive Web Response System (IWRS), when the patient first enters the system at screening. The randomization codes will be computer-generated by an [redacted] biostatistician and kept by an unblinded statistician [redacted] independent from the project team. The randomization will be done in blocks and stratified according to PFIC type (Type 1 and 2) and age category (6 months to 5 years, 6 to 12 years, and 13 to ≤18 years) to ensure approximate balance between dose schemes (1:1:1). A separate randomization list will be prepared for the maximum of 5 patients from Study A4250-003 regardless of stratification. Randomization codes will be assigned sequentially as patients become eligible for randomization.

Patients who are eligible for randomization will be assigned a unique 4-digit randomization number by IWRS. This randomization number identifies which treatment will be allocated to

the patient. Patients who withdraw from the study after the randomization visit will not be replaced and their randomization number will not be reused.

To ensure blinding, the study drug and the matching placebo have the same shape and size. Patients will receive capsule(s) of odevixibat according to their assigned dose group or capsule(s) of matching placebo once a day during the double-blind Treatment Period. Labels on the study drug containers will not identify the treatment to which a patient is randomized. Traceability of the treatment is ensured by the study drug number.

The IWRS system will assign a study drug number(s) corresponding to the randomization arm at each dispensing visit. The 5-digit study drug number will identify study drug packs and will be detailed on the study drug label. Dispensing of study drug will be coordinated by IWRS.

3.3 Study treatments and assessments

There will be three treatment groups in this study.

- Odevixibat 40 µg/kg/day group
- Odevixibat 120 µg/kg/day group
- Placebo group

Odevixibat and placebo will be supplied as capsules for oral administration. Study drug is to be taken once daily on Day 1 through Day 168 as described in Protocol Section 8.2. Study drug will be dispensed to the patient at Visits 3, 4, 5, 6, 7, and 8, together with instructions on how to store and take the drug. Study drug administration data, including whether each patient took each dose or partial doses of study drug, whether there were any delayed or missed doses, and whether the capsule was opened or swallowed whole will be documented through the eDiary and transferred to the clinical study database.

The number of capsules provided to the patient will be based on body weight at Visit 3. At Visit 6, if a patient's weight has shifted to another weight category (identified in Protocol Table 2), the number of capsules per day may be adjusted.

Odevixibat should be taken in the morning, together with food. On clinic visit days when laboratory assessments are conducted, study drug should be taken after the visit and after laboratory samples are taken. The last dose of study drug will be administered the day before Visit 9. At Study Visit 9, the patient has the option to enroll in an open-label extension study (A4250-008) should they meet the required entry criteria. If the patient is continuing into Study A4250-008 this will be the last visit in this study and the first visit in Study A4250-008. Visit 9/EOT should also be performed if the patient is prematurely withdrawn from the study.

Clinic Visit 10 will take place 28 days after Visit 9 for patients not continuing into Study A4250-008. A patient who withdraws from treatment prematurely and does not enter Study A4250-008 will have EOT (Visit 9) assessments at the time of withdrawal and a follow-up

assessment (Visit 10) 28 days following the last dose of study drug, and the patient will be recorded as withdrawn.

Efficacy assessments and daily recording of pruritus using an electronic diary (eDiary) in this study are briefly introduced as follows.

Serum Bile Acids

Blood samples for analysis of fasting total s-BA will be drawn at all visits. Samples will be processed and transported to a central laboratory per instructions in the laboratory manual. Patients will fast (water intake only is permissible) for at least 4 hours prior to the collection of samples for s-BA. Exceptions can be made for infants, <12 months of age, if unable to fast for the full 4 hours. For any visit at which a bile acid sample result is unreportable, an additional unscheduled visit for a repeat sample collection may be scheduled. All s-BA results during the treatment period and follow-up will be blinded.

Itching, Scratching and Sleep Score

Itching, observed scratching, and sleep disturbance will be recorded twice each day via eDiary. Patients and/or caregivers will be instructed to complete the eDiary every day throughout the study in the morning after the patient wakes up and, in the evening, just before the patient goes to sleep. Patients and caregivers will receive a formal training session during Visit 1.

The eDiary includes Albireo ObsRO and PRO items. Patients <8 years of age will not be asked to complete the Albireo PRO items; only the Albireo ObsRO will be completed by caregivers of patients in this age group. Older patients, 8 to 18 years of age, will complete the Albireo PRO items and the caregiver will complete the Albireo ObsRO items. The Albireo PRO items assess severity of itch, aspects of sleep disturbance (morning diary only), and tiredness. For patients 8 to 12 years of age, the caregiver will read the Albireo PRO items along with the child and record the child's response. A guide will be provided to the caregivers that provides standardized explanations of the Albireo PRO items, in case the patient is confused or requires clarification about the meaning of a question. The Albireo ObsRO items assess severity of observed scratching, aspects of observed sleep disturbance (morning diary only) and observed signs of tiredness (evening diary only). The Albireo ObsRO and PRO scratching and itch severity items use 0 to 4 response scales, where each response is distinguished by a unique facial expression, verbal anchor, number, and color code (Protocol Appendix 2).

Growth

Growth will be measured as height (velocity) and weight (Z-score) using a certified weight scale at Visits 1, 3, 6 and 9. BMI will be calculated by weight (kg) / height (m^2). Change will be defined as linear growth deficit (weight and BMI for age) compared to standard growth curve (Z-score, standard deviation from the 50^{th} percentile [SD from P50]).

Quality of Life Questionnaire (PedsQL)

Patients and caregivers will be asked to complete a QoL questionnaire (PedsQL) at Visit 3 and Visit 9. Details of the questions included on the questionnaire are provided in Protocol Appendix 4.

Biomarker

Blood samples for autotaxin and p-C4 will be drawn at Visits 3, 4, and 9 for children with body weight >10 kg.

PELD/MELD Score

At randomization and EOT, pediatric end-stage liver disease (PELD) score will be calculated for children under 12 years of age. For children 12 years or older, the (model for end-stage liver disease) MELD score will be calculated.

PELD Score = 4.80 * ln(total bilirubin) + 18.57 * ln(INR) - 6.87 * ln(albumin) + 4.36 (if patient <1 year: scores for patients listed for liver transplantation before the patient's first birthday continue to include the value assigned for age (<1 year) until the patient reaches the age of 24 months) + 6.67 (if the patient has growth failure [< -2SD])

Laboratory values less than 1.0 will be set to 1.0 for the calculation of the PELD score.

MELD Score = 9.57 * ln(creatinine) + 3.78 * ln (total bilirubin) + 11.2 *ln(INR) + 6.43
Laboratory values less than 1.0 will be set to 1.0 and serum creatinine values greater than 4.0 mg/dL (equivalent of 353.6 µmol/L) will be set to 4.0 for calculation of the MELD score.

For calculation of the PELD/MELD score [3] laboratory parameters will be converted (if necessary) into the following units:

- Total Bilirubin in mg/dL
- Albumin in g/dL
- Creatinine in mg/dL

Fibroscan®

Where available, Fibroscan® will be performed as per institution standard practice at Visits 3 and 9. The data collected on the Fibroscan form (liver stiffness measured in kPa and controlled attenuation parameter [CAP] measured in dB/m) will be converted to determine stage of fibrosis and grade of steatosis, respectively, using a score card [4,5] as outlined in the table below.

Fibroscan Scoring for Cholestatic Liver Disease (Fibrosis Score):

Score	F0 to F1 No scarring/ Mild fibrosis	F2 Moderate fibrosis	F3 Severe fibrosis	F4 Cirrhosis or Advanced fibrosis
FibroScan Result	2 to 7 kPa	>7 to 9 kPa	>9 to 17 kPa	>17 kPa

CAP Score & Steatosis Grading:

CAP Score	Amount of Liver with Fatty Change	Steatosis Grade
<238 dB/m		S0
238 to 260 dB/m	11 to 33%	S1
260 to 290 dB/m	34 to 66%	S2
> 290 dB/m	≥67%	S3

Markers of Fibrosis

APRI score and Fib-4 score will be calculated at randomization and at end of treatment.
APRI = [(AST in U/L) / (AST ULN in U/L)] × 100 / (Platelets in 10^9/L)
Fibrosis 4 Score = (Age*AST in U/L) / (Platelets in 10^9/L * √ (ALT in U/L)).

Global Impression of Change and Global Impression of Symptom Measures

Patients ≥8 years of age, caregivers, and clinicians will complete the global impression of change (GIC) and the global impression of symptoms (GIS) measures at randomization (Visit 3; GIS only), Visits 4 and 6, and at EOT (Visit 9). The GIC items assess change in itch (patient version), scratching (caregiver and clinician versions), and sleep (all versions) since starting the study drug. The GIS items assess itch (patient version), scratching (caregiver and clinician versions), and sleep (all versions) in the past week.

A detailed description of procedures and assessments to be conducted during this study is summarized in the schedule of study assessments in Table 1 below.

Table 1: Schedule of Study Assessments

Study Activity	Screening Period		Treatment Period								Follow-Up
			Random-ization	Phone Contact	4 weeks	8 weeks	12 weeks	18 weeks	22 weeks	24 weeks	
Study Days	**-56 – (-35) ±2**	**-28 – (-7) ±2**	**0**	**14 ±3**	**28 ±3**	**56±3**	**84±3**	**126 ±3**	**154 ±3**	**168 ±3**	**196 ±3**
Visits	**Clinic Visit 1**	**Clinic Visit 2**	**Clinic Visit 3**	**n/a**	**Clinic Visit 4**	**Clinic Visit 5**	**Clinic Visit 6**	**Clinic Visit 7**	**Clinic Visit 8**	**Clinic Visit 9/EOT[a]**	**Clinic Visit 10[b]**
Informed consent[c]	X										
Inclusion/exclusion criteria	X		X								
Demographics	X										
Concomitant medication[d]	X	X	X	X	X	X	X	X	X	X	X
Medical and surgical history	X		X								
Physical examination including voluntary photography	X		X				X			X	
Skin examination	X		X		X		X			X	X
Vital signs[e]	X	X	X		X	X	X	X	X	X	X
eDiary: itching, scratching, and sleep scores[f]	Daily diary entry										
Clinical chemistry[g]	X		X		X	X	X	X		X	X
Hematology[g]			X		X	X	X	X		X	X
Urinalysis[g]		X					X			X	
Serum bile acids[h]	X	X	X		X	X	X	X	X	X	X
INR		X			X		X		X		X
Autotaxin[i]			X		X					X	

Study Activity	Screening Period		Treatment Period								Follow-Up
			Random-ization	Phone Contact	4 weeks	8 weeks	12 weeks	18 weeks	22 weeks	24 weeks	
Study Days	**-56 – (-35) ±2**	**-28 – (-7) ±2**	**0**	**14 ±3**	**28 ±3**	**56±3**	**84±3**	**126 ±3**	**154 ±3**	**168 ±3**	**196 ±3**
Visits	**Clinic Visit 1**	**Clinic Visit 2**	**Clinic Visit 3**	**n/a**	**Clinic Visit 4**	**Clinic Visit 5**	**Clinic Visit 6**	**Clinic Visit 7**	**Clinic Visit 8**	**Clinic Visit 9/EOT[a]**	**Clinic Visit 10[b]**
p-C4[i]			X		X					X	
AFP			X							X	
Vitamins A & E[h]		X					X			X	
25-hydroxy vitamin D			X				X			X	
Blood sample for A4250 PK[i]					X					X	
Abdominal Ultrasound			X							X	
Fibroscan® (where available)			X							X	
QoL questionnaire (PedsQL)			X							X	
Patient/Caregiver/Clinician Global Impression of Change					X		X			X	
Patient/Caregiver/Clinician Global Impression of Symptoms			X		X		X			X	
Pregnancy test[j]	X[j]	X	X		X	X	X	X	X	X[j]	X
Study drug dispensed[k]			X		X	X	X	X	X		
Adverse events[l]	Continuous collection										
Study drug compliance					X	X	X	X	X	X	

AFP: alfa-fetoprotein; BMI: body mass index; eCRF: electronic case report form; eDiary: electronic diary; EOT: End of Treatment; ICF: informed consent form; INR: international normalized ratio; p-C4: plasma 7α Hydroxy 4-cholesten-3-one; PedsQL: Pediatric Quality of Life Inventory; QoL: quality of life; RR: respiratory rate; s-BA: serum bile acid; PK: pharmacokinetics.

a Must also be performed at the time a patient is prematurely withdrawn from the study. This visit is the same as Visit 1 in the extension protocol (Study A4250-008).

b Assessments must be performed 28 days following the final dose of study drug. This visit is omitted if patients are entering Study A4250-008.

c The ICF (and assent) must be signed before any study procedures are performed.

d Past (within 1 month prior to Visit 1) and current medication will be documented in the eCRF.

e Includes blood pressure, pulse, RR, temperature, height (or length depending on age), and body weight (using a certified weight scale). BMI will be calculated.

f Includes training on use of the device at Screening (Visit 1) and review of patient daily entries during 14 days prior to Clinic Visit 3 to determine eligibility.

g See protocol Table 3 for detailed parameters.

h Patients will fast (water intake only is permissible) for at least 4 hours prior to the collection of samples for s-BA and vitamin A. Exceptions can be made for infants, less than 12 months of age, if unable to fast for the full 4 hours. All s-BA results during the treatment period and follow-up will be blinded.

i Samples will not be collected for patients ≤10 kg. In the event of an SAE, a PK sample should be collected as close to the onset of the event as possible.

j For girls who have reached menarche. Serum test at Visit 1; urine test at other visits. A serum pregnancy test will also be conducted at Visit 9 for patients entering Study A4250-008. If a urine pregnancy test is positive, a serum pregnancy test should be performed to confirm the pregnancy.

k Study drug to be taken once daily on Day 1 through Day 167 as described in protocol Section 8.2

l Adverse event information will be collected from the time of signing of the ICF to study discontinuation.

4 STUDY ENDPOINTS

4.1 Primary efficacy endpoints

The primary efficacy endpoints are as follows:

- **Europe (EU) and rest of the world (RoW)**: Proportion of patients experiencing at least a 70% reduction in fasting s-BA concentration from baseline to the end of treatment or reaching a level ≤70 µmol/L after 24 weeks of treatment.

  Fasting s-BA baseline will be calculated as the average of the last 2 values prior to the first dose. The end value is the average of the values at Weeks 22 and 24 after the start of double-blind treatment. If a patient's baseline value is ≤70 µmol/L, then only the criterion of at least 70% reduction in fasting s-BA concentration will be used to determine whether the patient is a responder or not for the primary endpoint analysis.

- **United States (US):** Proportion of positive pruritus assessments at the subject level over the 24-week treatment period based on the Albireo ObsRO instrument.

  A positive pruritus assessment is defined as a scratching score of ≤1 or at least a one-point drop from baseline on the Albireo ObsRO instrument. At each assessment, the AM score will be compared to the baseline AM average, and the PM score will be compared to the baseline PM average. Both AM and PM pruritus assessments will be included in the analysis of this endpoint. AM scores from the period of 14 days prior to or on the first dose day of study medication will be averaged as AM baseline. PM scores from the period of 14 days prior to the first dose day of study medication will be averaged as PM baseline. If a patient's baseline average score is ≤ 1, then only the criterion of a one-point drop from baseline on the Albireo ObsRO instrument will be used to determine whether a pruritus assessment is positive or not for the primary endpoint analysis.

4.2 Secondary efficacy endpoints

The secondary efficacy endpoints include the following:

- **EU and RoW**: Proportion of positive pruritus assessments at the subject level over the 24-week Treatment Period based on the Albireo ObsRO instrument.

- **US**: Proportion of patients experiencing at least a 70% reduction in fasting s-BA concentration from baseline to the end of treatment or reaching a level ≤70 µmol/L compared to placebo after 24 weeks of treatment.

- **All Regions**:

  All other secondary endpoints are compared to placebo.

  - Change from baseline to Week 12 and to Week 24 in fasting s-BA
  - Change from baseline to Week 12 and to Week 24 in serum ALT concentration
  - Change in growth from baseline to Week 12 and to Week 24, defined as the linear deficit (height for age, weight for age and MBI for age) compared to a standard growth curve (Z-score, SD from P50)
  - Proportion of responders for pruritus scores at Weeks 12 and 24 based on the ObsRO instrument. For pruritus scores, a responder is defined as a patient achieving a reduction in score equivalent to or exceeding the clinically meaningful threshold established from the blinded psychometric analysis (Pruritus Evidence Dossier).
  - Change in sleep parameters measured with the Albireo PRO and ObsRO instruments by each 4-week interval over the 24-week treatment period
  - Proportion of individual assessments meeting the definition of a positive pruritus assessment at the subject level over the 24-week treatment period as reported on the Albireo PRO instrument. A positive pruritus assessment includes an itch score ≤1, or at least a one-point drop from baseline based on the Albireo PRO instrument; only patients ≥8 years of age will complete the Albireo PRO instrument
  - Proportion of individual assessments meeting the definition of a positive pruritus assessment at the subject level from Weeks 0-4, Weeks 0-8, Weeks 0-12, Weeks 0-18, Weeks 0-24, respectively, and the proportion of positive pruritus assessments at each 4-week interval as reported on the Albireo ObsRO instrument. A positive pruritus assessment is defined as a scratching score of ≤1 or at least a one-point drop from baseline on the Albireo ObsRO instrument. At each assessment, the AM or PM score will be compared to the baseline AM or PM average, respectively.
  - Proportion of individual AM assessments meeting the definition of a positive pruritus assessment at the subject level from Weeks 0-4, Weeks 0-8, Weeks 0-12, Weeks 0-18, Weeks 0-24, respectively, and the proportion of positive pruritus assessments at each 4-week interval as reported on the Albireo ObsRO instrument. A positive pruritus assessment is defined as a scratching score of ≤1 or at least a one-point drop from baseline on the Albireo ObsRO instrument. At each assessment, the AM score will be compared to the baseline AM average.
  - Proportion of individual PM assessments meeting the definition of a positive pruritus assessment at the subject level from Weeks 0-4, Weeks 0-8, Weeks 0-12,

Weeks 0-18, Weeks 0-24, respectively, and the proportion of positive pruritus assessments at each 4-week interval as reported on the Albireo ObsRO instrument. A positive pruritus assessment is defined as a scratching score of ≤1 or at least a one-point drop from baseline on the Albireo ObsRO instrument. At each assessment, the PM score will be compared to the baseline PM average.

- Number of patients undergoing biliary diversion surgery or liver transplantation
- Number and percent of patients achieving positive pruritus assessment for more than 50% of the time during the 24-week treatment period

4.3 Exploratory endpoints

Exploratory efficacy endpoints include the following:

- Assessment of Global Symptom Relief at Weeks 4, 12, and 24 as measured by patient, caregiver, and clinician via the global impression of change (GIC) item
- Change from baseline in fasting s-BA at Week 4
- Change in pruritus, as indexed by caregiver-reported observed scratching, from baseline by each 4-week interval of the Treatment Period as measured by the Albireo ObsRO instrument
- Change from baseline by each 4-week interval of the Treatment Period in patient-reported and observer-reported night time itching and scratching severity scores, respectively, as measured by the Albireo PRO and ObsRO instruments.
- Change from baseline by each 4-week interval of the Treatment Period in patient-reported and observer-reported morning time itching and scratching severity scores, respectively, as measured by the Albireo PRO and ObsRO instruments.
- Change from baseline by each 4-week interval of the Treatment Period in pooled pruritus score including observer-reported scratching for patients <8 years of age and patient-reported itch severity for patients ≥8 years of age, as measured by the Albireo PRO and ObsRO instruments, respectively.
- Change from baseline by each 4-week interval of the Treatment Period in additional patient-reported and observer-reported sleep parameters (e.g. tiredness and number of awakenings), as measured by the Albireo PRO and ObsRO instruments, respectively.
- Change from baseline to Week 24 in PedsQL
- Change from baseline in serum ALT concentration at Week 4

- Change from baseline in total bilirubin at Weeks 4, 12, and 24
- Change from baseline in AST and gamma-glutamyl transferase at Weeks 4, 12, and 24
- Change from baseline for other biochemical markers and measures of bile acid synthesis (autotaxin, plasma 7α-hydroxy-4-cholesten-3-one concentration [p-C4]) at Week 24
- Change in paediatric end-stage liver disease/model for end-stage liver disease (PELD/MELD) score, AST to platelet ratio index (APRI) score, and Fib-4 score from baseline to Week 24
- Change from baseline to Week 24 in stage of liver fibrosis as assessed by Fibroscan® (where available)
- For patients who answered Yes to Questions 1 and 2 in the exit survey, changes from baseline in the average scratching score based on the Albireo ObsRO instrument in the last 2 weeks before treatment termination
- For patients who answered No to Questions 1 or 2 in the exit survey, changes from baseline in the average scratching score based on the Albireo ObsRO instrument in the last 2 weeks before treatment termination

4.4 Safety

Safety criteria are as follows:

- Primary safety analysis is the incidence of treatment-emergent adverse events (TEAEs) and TEAEs categorized by causality, severity, and seriousness assessments made by the investigator by comparing active arms vs. placebo. This includes liver-related mortality and liver decompensation events (protocol Section 10.2.2.3) and all-cause mortality.
- Safety will also be evaluated by the following assessments:

  - Physical examinations
  - Concomitant medications
  - Vital signs
  - Laboratory test results (including clinical chemistry, hematology, urinalysis, AFP, vitamins A and E, 25-hydroxy vitamin D, and INR)
  - Abdominal ultrasound
  - Discontinuations due to AEs

5 SAMPLE SIZE AND POWER

For the EU and RoW, the primary analysis will be related to the serum bile acid responder endpoint. For the US, the primary analysis will be related to the pruritus endpoint. The CMH test stratified by randomization strata will be used to test the serum bile acid responder endpoint. The individual null hypothesis is that the odds ratio of the response in an active arm vs. placebo is 1. The alternative hypothesis is that the odds ratio is greater than 1. An ANCOVA will be used for treatment comparisons for the proportion of positive pruritus assessments at subject level. The individual null hypothesis is that the average proportion of positive assessments is the same between the active and placebo arms. The alternative hypothesis is that the average proportion is larger in the active arm.

Since each active arm will be compared with the placebo, there will be two individual null hypotheses for each endpoint. In each endpoint, a pooled analysis for the closed testing procedure will be applied to control the 1-sided overall type I error rate for the two odevixibat treatment comparisons vs. placebo at the 0.025 level [6,7,8]. The significance of an active arm will be claimed if both 1-sided p-values in the pooled analysis and individual treatment comparison are less than or equal to 0.025.

The study will enroll 60 to 70 patients in order to obtain at least 20 evaluable patients in each arm. For each primary endpoint, simulations with 5,000 iterations using 20 patients per arm were conducted to estimate the power after multiplicity adjustment, resulting in a standard error (SE) of <0.7% for each estimated power.

Based on the Phase 2 study (A4250-003) data, both low and high dose groups are assumed to have the same positive treatment effects in both the serum bile acid and pruritus endpoints in the simulation. For serum bile acids, binomial distributions were used to simulate the proportion of responders to estimate the power. The simulated proportions were analyzed using the Cochran-Mantel-Haenszel (CMH) test to generate the following 1-sided p-values: both odevixibat arms pooled vs. placebo, low dose vs. placebo, and high dose vs. placebo. Assuming 60% responders in the odevixibat arms and 10% responders in the placebo arm, the power to claim significance for a particular odevixibat arm after multiplicity adjustment is approximately 94%. The probability to claim significance for at least one arm, and for both arms are approximately 99% and 91%, respectively. If the response rates are 50% in the odevixibat arms and 10% in the placebo arm, with 20 patients per arm, the probability to claim significance for a particular odevixibat arm after multiplicity adjustment is approximately 82%. The probability to claim significance for at least one arm, and for both arms is approximately 91% and 73%, respectively.

For the proportion of positive pruritus assessments at the subject level in pruritus scores, beta-binomial distributions were used for power simulations since the binary results of the individual assessments within a subject are likely correlated [9]. The effect size was 1.0526 from the original sample size calculation using change from baseline as the endpoint. The same effect size was assumed for the current endpoint for the low and high dose vs. the control. A difference of 15%, 20%, 25%, and 30% in proportion of positive pruritus assessments were considered in power simulation. Within each difference, proportions of positive assessments in the placebo arm ranging from 15% to 35% were considered. Subsequently, the proportion of positive

assessments in an active arm, the standard deviation, and the corresponding beta binomial parameters were calculated to satisfy the assumed effect size. These parameters were used to simulate correlated binary results within each subject. The simulated proportions were analyzed using analysis of variance (ANOVA) to generate the following 1-sided p-values: both odevixibat arms pooled vs. placebo, low dose vs. placebo, and high dose vs. placebo. The simulation in each scenario was repeated for 5,000 iterations using the current sample size of 20 patients per arm. The simulated power to claim significance for a particular arm after multiplicity adjustment is quite consistent under different scenarios and is approximately 89%. The probability to claim significance for at least one arm, and for both arms is approximately 95% and 83%, respectively.

6 ANALYSIS SETS

6.1 Full Analysis Set (FAS)

The full analysis set (FAS) will consist of all randomized patients who received at least 1 dose of study treatment. Patients will be analyzed in the treatment group they were randomized to even if they received incorrect study drug. The FAS will be the primary analysis set for efficacy analyses.

6.2 Safety Analysis Set (SAS)

The safety analysis set will consist of all randomized patients who received at least 1 dose of study drug. Patients will be analyzed according to the treatment they actually received. In case of wrong study drug dispensed to a patient in error, the patient will be assigned to the treatment arm that the actual drug was mostly taken. The safety analysis set will be used for safety analyses.

6.3 Per-Protocol Analysis Set (PP)

The per protocol (PP) analysis set is a subset of the FAS and will consist of all randomized patients for whom no important protocol deviation (no deviation which may affect the study efficacy outcome) is documented. In identifying this population, important protocol deviations are defined as:

- Failure to meet the following inclusion criteria
    - Inclusion criteria (2) Patient must have a clinical genetic confirmation of PFIC-1 or PFIC-2 through identification of biallelic pathogenic variants in either the *ATP8B1* or *ABCB11* genes
    - Inclusion criteria (3) Patient must have elevated s-BA concentration, specifically measured to be ≥100 µmol/L, taken as the average of 2 samples at least 7 days apart (Visits 1 and 2) prior to randomization
    - Inclusion criteria (4) Patient must have history of significant pruritus and a caregiver-reported observed scratching in the eDiary average of ≥2 (on 0 to 4 scale) in the 2 weeks prior to randomization
- Enrollment in spite of meeting the following exclusion criteria
    - Exclusion criteria (1) Patient with pathologic variations of the *ABCB11* gene that predict complete absence of the BSEP protein
    - Exclusion criteria (15) Patient suffers from uncontrolled, recalcitrant pruritic condition other than PFIC. Examples include, but not limited to, refractory atopic dermatitis or other primary pruritic skin diseases
- Compliance rate of <80% with study drug administration estimated based on CRF data

Important protocol deviations will be also identified by study team during protocol deviation review. The PP analysis set will provide supportive data for the efficacy analyses of primary and selected secondary endpoints.

7 STATISTICAL CONSIDERATIONS AND ANALYSIS

7.1 Handling of missing data and/or invalid data and outliers

7.1.1 Handling of Missing data for efficacy endpoints

7.1.1.1 Pruritus Endpoint

In the primary analysis of the pruritus endpoint, all intermittently missing assessments will be classified as negative pruritus assessments (that is, not meeting the definition of a positive assessment as defined in Section 4.1). Similarly, all missing planned assessments after premature treatment discontinuation will be counted as negative pruritus assessments. All planned assessments after death or initiation of rescue treatments such as biliary diversion surgery or liver transplantation will be counted as negative pruritus assessments. Both on-treatment and off-treatment pruritus assessments up to Day 168 will be included for analysis.

In general, all 336 (=168 x 2) assessments planned over the 24-week treatment period will be included in the denominator of the endpoint definition. However, if the last pruritus assessment for a patient who completes treatment (i.e., completes Week 24 assessment) is reported prior to Day 168, then the denominator used for analysis will be (the last report day x 2). The last report day will be determined for ObsRO and PRO instruments separately.

For the analysis of the pruritus endpoint, an additional method will also be used to calculate the proportion of positive pruritus assessments up to the last report day before or on Day 168. In other words, the denominator to calculate the proportion will be (the last report day x 2). All missing data before or on the last report day will be considered as negative assessments.

In addition, a multiple imputation method will be performed for pruritus assessments as a continuous endpoint by considering the impact of intermittently missing data as well as intercurrent events (Protocol Section 11.2.2.2). Moreover, a tipping point analysis will be performed to further evaluate the influence of missing data. The details of this are addressed in the Section 8.6.3.1.

7.1.1.2 Serum bile acid responder endpoint

In the primary analysis of the fasting bile acid responder endpoint, dropouts and treatment completers with a missing average at the end of the treatment will be treated as non-responders. Patients in need of surgical rescue (i.e., biliary diversion and/or liver transplant) will be classified as non-responders.

As a sensitivity analysis, a tipping point analysis of fasting bile acid responder endpoint will be performed to evaluate the strength of evidence if the result is significant. The details of this are addressed in the Section 8.6.3.1.

8 STATISTICAL METHODS

8.1 General statistical conventions

All personnel involved with the analysis of the study will remain blinded until database lock. For DSMB analyses see Section 8.9.2. All statistical procedures will be completed using SAS® version 9.4 or higher.

Statistical testing for each primary endpoint analysis will be performed with a 1-sided overall Type I error rate of 0.025, adjusting for 2 comparisons between each active arm vs. the control. All other secondary endpoints for treatment comparisons will not be adjusted for multiplicity and nominal p values will be provided. Two-sided 95% confidence intervals (CI) will be provided when relevant.

For analysis adjusting for stratification factors (PFIC type and age category), if a patient is under 6 months due to imputed birth date (Appendix A), then the patient will be grouped to the age category '6 months to 5 years'.

Continuous variables will be summarized using descriptive statistics, including the number of patients with non-missing value (n), mean, median, SD, minimum and maximum. "n" will be presented without a decimal point, minimum and maximum values will be presented in the same precision as in the database, mean and median will be presented in one more decimal place than the minimum and maximum, and SD will be presented in one more decimal place than the mean and median.

For categorical variables, summaries will include counts of patients (frequencies) and percentages. Percentages will be rounded to one decimal place. Descriptive summaries of change from baseline in categorical variables will be provided using shift tables, if applicable.

For summary purposes, if not otherwise specified, the baseline value of a parameter is defined as the last non-missing assessment of that parameter prior to the first intake of study drug (Appendix A). Derived variables used for the analyses are provided in Appendix A. All summaries will be presented by treatment group, unless otherwise specified.

For summary tables, the following 2 presentations will be used:

- For baseline related tables and protocol deviations: odevixibat 40 µg/kg/day, odevixibat 120 µg/kg/day, odevixibat 2 doses combined, placebo, and overall (3 treatment groups combined) will be presented.
- For the safety and efficacy related tables: odevixibat 40 µg/kg/day, odevixibat 120 µg/kg/day, odevixibat 2 doses combined, and placebo will be presented.

The visit window details for the derivation of study weeks and time points are outlined in Appendix B.

All patient data, including those derived, will be presented in individual patient data listings. All

listings will be sorted by patient ID, treatment group, date/time and visit. Each patient's sex and age will be stated on each listing. Data listings will be based on all randomized patients.

8.2 Subject disposition

All patients who provided informed consent will be included in a summary of patient accountability.

The following categories will be summarized (by treatment group and overall where applicable):

- Patients screened
- Screening failures
- Patients randomized
- Patients randomized from Study A4250-003
- Patients randomized and dosed
- Patients completing treatment
- Patients completing the study
- Patients discontinuing treatment (including reasons for treatment discontinuation)
- Patients withdrawing early from study (including withdrawal reasons)
- Patients entering the extension study (A4250-008)

Additionally, patients included in the FAS analysis set will be summarized by region (Appendix C).

8.3 Protocol deviations

Important protocol deviations will be identified prior to database lock. All exclusions of patients from the per-protocol (PP) analysis sets will be performed programmatically using the criteria of important protocol deviations (Section 6.3) in a blinded way and reviewed at the final protocol deviations review meeting prior to the final database lock, through clinical input provided by the sponsor, using the following sources of information:

- Supportive subject listings, provided by the [redacted] lead statistician ahead of the protocol deviation review meeting, based on data recorded on the eCRF.
- Protocol deviation logs, provided by [redacted] Clinical, Pharmacovigilance, Medical and Data Management.

The number of patients with identified important protocol deviations will be summarized by treatment group and overall (on randomized patients). Details on important protocol deviations will be provided in a listing on all randomized patients.

In addition, a listing will be provided with all protocol deviations identified based on data recorded on the eCRF and/or protocol deviation logs from [redacted] Medical and Data Management (on randomized patients) prior to database lock and documented in the CSR.

8.4 Demographics and baseline characteristics

8.4.1 Demographics

Age, height, weight and BMI at baseline and other demographic variables e.g., age category (6 months to 5 years, 6 to 12 years, and 13 to 18 years), sex, race, ethnicity, country, and region (see Appendix C) will be summarized by treatment group and overall descriptively using the FAS analysis set.

8.4.2 Baseline and disease characteristics

The following disease characteristics will be summarized by treatment group and overall: years since PFIC diagnosis, type of PFIC, presence of significant pruritus per investigator report, at least one s-BA level >100 umol/L within 6 months prior to screening visit, reason for discontinuation of historical PFIC related investigational medications, and diagnostic genetic laboratory test as well as Child-Pugh classification and hepatic impairment classification per NCI Organ Dysfunction Working Group for the FAS analysis set.

Child-Pugh classification: mild (Class A), moderate (Class B), or severe hepatic impairment (Class C) is based on the FDA Guidance for Pharmacokinetics in Patients with Impaired Hepatic Function. Determination of the classification is made based on medical review of baseline laboratory data and medical history.

Hepatic impairment will also be classified per the NCI Organ Dysfunction Working Group classification as mild, moderate or severe (Mansfield et al. 2016 [10]). Both the Child-Pugh and NCI classification will be included in the baseline characteristics and both will be applied in the subgroup analysis based on hepatic impairment status.

The following continuous parameters will also be summarized: estimated glomerular filtration rate (eGFR); baseline Z-score (weight, height, and BMI); and baseline values of alanine aminotransferase (ALT); aspartate aminotransferase (AST); gamma-glutamyl transferase (GGT); total bilirubin; alkaline phosphatase; INR; and vitamins A, E, and 25 hydroxy D.

eGFR will be calculated based on the modified/bedside Schwartz equation. For patients <18 year of age, Bedside Schwartz Equation will be used:
GFR (mL/min/1.73 m²) = (36.2 × Height in cm) / Creatinine in µmol/L

For patients ≥ 18 years, the isotope dilution mass spectrometry (IDMS)-traceable Modification of Diet in Renal Disease (MDRD) Study equation will be used:
GFR (mL/min/1.73 m^2) = 175 × (Creatinine in µmol/L/88.4)$^{-1.154}$ × (Age)$^{-0.203}$ × (0.742 if female) × (1.212 if African American)

A listing will be created to present all the demographic and baseline characteristics for the FAS analysis set. In addition, an individual listing on liver biopsies performed within 1 year prior to screening will be produced.

The results of pathologic variants identified for ATP8B1 or ABCB11 genes will be listed for all patients for the FAS analysis set.

8.4.3 Medical and surgical history

The frequencies and percentages of patients with reported medical and surgical history will be presented by system organ class (SOC) and preferred term (PT) by treatment group and overall for the SAF analysis set. Medical and surgical history will be coded using MedDRA version 23.0 throughout the study. The summary table will be sorted alphabetically for SOC and PT. Medical and surgical history will also be listed.

8.4.4 Prior and concomitant medications

Prior and concomitant medications used in this study will be coded using the version of the World Health Organization Drug Global Dictionary March 2020 throughout the study.

Prior medications: All medications taken by a patient within 3 months prior to the first intake of study drug with a recorded medication stop date prior to the first intake of study drug are regarded as prior medication. This will not include historical PFIC related investigational medications collected separately on CRFs.

Concomitant medications: All medications taken by a patient on or after the first intake of study drug and medications with start date before the first intake of study drug without a recorded stop date prior to the first intake of study drug.

Prior and concomitant medication use during the study will be summarized descriptively using frequency tables by Anatomical Therapeutic Chemical (ATC) class 4 and preferred term (PT) by treatment group for the SAF analysis set. ATC class and PT will be presented alphabetically. All prior and concomitant medication will be listed. Details for imputing missing or partial start and/or stop dates of non-study medication are described in Appendix D.

8.5 Extent of exposure

8.5.1 Treatment duration

Exposure will be summarized with descriptive statistics (n, mean, SD, minimum, median, and maximum) and presented by treatment group using the SAF analysis set. The exposure of study drug will be derived as follows:

Duration of exposure (in weeks) = (Date of last study drug intake – Date of first study drug intake + 1)/7.

Investigators are allowed to interrupt the study drug to allow adverse events to resolve, if necessary. Any drug interruption will not be considered when calculating the treatment duration.

8.5.2 Study drug compliance

Study drug compliance will be assessed by maintaining adequate study drug dispensing records. The study drug compliance over the treatment period will be calculated using 2 sources of data.

Source Data from CRFs

Treatment Compliance = 100 × [(Number of capsules dispensed – number of capsules returned)/number of capsules that should have been taken].

The number of capsules that should have been taken is calculated as the number of days that the patient was in the treatment period (exposure as above) multiplied by the number of prescribed capsules to be taken (based on patient's body weight, protocol Table 2) during the treatment period. The total number of capsules actually taken is the total number of capsules recorded as taken based on the CRF (number of capsules dispensed minus returned) summed over the treatment period. If the number of capsules not returned is confirmed as missing and the study drug is confirmed as not having been returned, the derivation will not be done.

Source Data from eDiary

Treatment Compliance =100 x (total number of capsules taken/total number of capsules planned to be taken). The number of capsules planned to be taken will be estimated based on patient's body weight at Visit 3 and Visit 6 per protocol Table 2. The number of capsules planned to be taken before Visit 6 will be the same as the number estimated at Visit 3, and after Visit 6 will be the same as the number estimated at Visit 6.

The calculated compliance rates will be summarized and the one based on CRFs data will be considered the primary compliance rate. The compliance rates reported by visit on the CRFs will be summarized also.

Descriptive summary statistics will be used to summarize study drug compliance by treatment group. The number and percentage of patients with compliance <80%, between 80% and 120%, and >120% will also be presented based on the SAF analysis set.

8.6 Efficacy analyses

This section addresses the analyses to be conducted on the primary, secondary, and exploratory efficacy variables. The primary objective of this study is to demonstrate the efficacy of repeated daily doses of 40 µg/kg/day and 120 µg/kg/day of odevixibat in children with progressive familial intrahepatic cholestasis (PFIC) Types 1 and 2, as determined by the following:

• Proportion of patients experiencing at least a 70% reduction in fasting serum bile acid (s-BA)

concentration from baseline to end of treatment or reaching a level ≤70 µmol/L compared to placebo after 24 weeks of treatment

• The proportion of positive pruritus assessments at the subject level over the 24-week Treatment Period

The efficacy analysis will be carried out using the patients from the FAS and/or PP analysis set, unless otherwise specified. Figures for selected efficacy endpoints, such as the primary pruritus endpoint, change in s-BA, change in other patient reported and caregiver reported outcomes, will be provided by treatment group and by visit.

For efficacy endpoints derived from central laboratory data, if central laboratory data are not available due to COVID-19 or other reasons, local laboratory data will be used.

8.6.1 Analyses methods

8.6.1.1 Cochran Mantel Haenszel (CMH) Test

The CMH test is used in the analysis of stratified categorical data. It tests the association between treatment and a binary outcome such as responder while considering the stratification factors. For the proportion of patients experiencing at least a 70% reduction in fasting s-BA concentration from baseline to the end of treatment or reaching a level ≤70 µmol/L after 24 weeks of treatment, the CMH test will be used for treatment comparisons. CMH stratified by PFIC type and age category will be performed to compare the proportion of the responders in fasting bile acid at the end of treatment (based on average of Weeks 22 and 24). The testing procedure is addressed in the Section 8.6.1.3.

In the CMH test, data in a stratum will not be used in the calculation of the p-value if a row sum or column sum is 0 in the contingency table. To ensure that all data are used when this occurs, data will be pooled with the adjacent age stratum that has a smaller number of subjects. If a row sum or column sum is still 0 after pooling data, all age groups will be pooled (i.e. the CMH test will be stratified by PFIC type only). If a row sum or column sum is still 0, the CMH test will not be stratified. Pooling strategy will also be conducted for a stratum with <4 patients.

The proportion together with corresponding Clopper-Pearson exact 95% CI, odds ratio and corresponding 95% CI, and p-value for the CMH test will be presented. The proportion difference with corresponding exact unconditional 95% CI without adjusting stratification factors will be presented. Miettinen-Nurminen (score) CI with adjusting stratification factors will be reported for common risk difference (i.e., proportion difference). The exact CI will be reported for the common odds ratio by using an algorithm based on Vollset, Hirji, and Elashoff (1991) [**11**]. SAS code will be found in Appendix E.

8.6.1.2 Analysis of Covariance (ANCOVA) model

The ANCOVA model is usually used to test the main and interaction effects of categorical variables on a continuous dependent variable, controlling for the covariates. The ANCOVA model will be used to analyze the comparisons of the proportion of positive pruritus assessments

at subject level over the 24-week treatment period between the treatment groups.

8.6.1.3 Multiplicity

For EU and RoW, the primary analysis will be related to the serum bile acid responder endpoint. For US, the primary analysis will be related to the pruritus endpoint. For each primary endpoint, a pooled analysis for the closed testing procedure will be applied to control the 1-sided overall type I error rate for two treatment comparisons vs. the placebo at the 0.025 level as specified below:

In the closed testing procedure, the low and high dose groups are pooled to compare with the placebo group first. If the 1-sided p-value is ≤0.025, the 1-sided p-values for low dose vs. placebo and high dose vs. placebo, respectively will be calculated. If both individual p-values are ≤0.025, a significant treatment effect will be declared on both dose groups. If only one of them is ≤0.025, a significant treatment effect will be declared on the corresponding dose group.

In testing the primary endpoint for US, treatment arms will be separated in the pooled analysis and the “lsmestimate" statement in SAS Proc Mixed will be used to compare pooled active arms vs. the placebo. See Appendix E.

Adjusted p-values will also be reported when unadjusted p-values are presented directly from the model. The adjusted p-value for an individual dose is calculated as the maximum value of the unadjusted p-value for the pooled low and high doses and the unadjusted p-value for the individual doses.

Analyses of secondary and exploratory endpoints will provide supportive efficacy and safety information regarding the differences between the treatment groups. No adjustments will be performed for multiple comparisons when testing these secondary and exploratory endpoints.

8.6.2 Treatment by center interaction analysis (multi-center study)

There will be only a few patients per center, so no direct exploration of treatment-by-center interaction effect is planned. Additional treatment-by-covariate interactions may be explored.

8.6.3 Analyses of primary efficacy endpoints

The primary analyses on the primary endpoints will be based on the FAS. Additional analyses on the primary endpoints will be conducted for PP analysis set.

For EU and RoW, proportion of patients experiencing at least a 70% reduction in fasting s-BA concentration from baseline to the end of treatment or reaching a level ≤70 µmol/L after 24 weeks of treatment will be analyzed using the CMH test. Patients with missing data at the end of treatment will be classified as non-responders. CMH stratified by PFIC type and age category will be performed to compare the proportion of the responders in fasting bile acid at the end of treatment (based on average of Weeks 22 and 24). The testing procedure is addressed in the Section 8.6.1.3. The contrasts of primary interest will be the treatment differences at end of treatment (Week 22 and 24); also, the difference at Week 4, Week 8, Week 12, and Week 18 will

be estimated and analyzed.

For US, the primary endpoint is the proportion of positive pruritus assessments at the subject level over the 24-week Treatment Period. A positive pruritus assessment is defined as a scratching score of ≤1 or at least a one-point drop from baseline on the Albireo ObsRO instrument. At each assessment, the AM score will be compared to the baseline AM average, and the PM score will be compared to the baseline PM average. Both AM and PM pruritus assessments will be included in the analysis of this endpoint. For analysis purposes, diary entries will be assigned to a study day based on the recorded date regardless of recorded time. AM scores from the period of 14 days prior to or on the first dose day of study medication will be averaged as AM baseline. PM scores from the period of 14 days prior to the first dose day of study medication will be averaged as PM baseline. Baseline score will be rounded to an integer to evaluate the positive pruritus assessments for the primary analysis. Additional analysis will be conducted using unrounded baseline score.

An ANCOVA model will be used for treatment comparisons. The model will include treatment arm, AM baseline pruritus score, PM baseline pruritus score, and stratification factors, i.e., PFIC type and age category. The testing procedure is addressed in the section 8.6.1.3.

LS mean (SE) by treatment arm and LS mean difference (SE), 95% confidence intervals, and p-values (where applicable) between treatments vs. placebo will be provided (see Section 8.6.1.3 below). SAS code will be found in Appendix E. If there are concerns on model assumptions, normality will be checked based on Shapiro-Wilk test and homogeneity of variances will be checked based on Levene's test. For normality testing, Shapiro-Wilk test will be performed for each arm. The p-value from each arm will be combined to get an overall p-value based on Fisher's combined probability test.

A non-parametric ANCOVA based on the rank scores will be used as a sensitivity analysis to calculate the p-values. The proportion of positive pruritus assessments at subject level over 24-week treatment period will be ranked ignoring treatment group. The value of midranks will be used in the case of ties. The baseline pruritus scores will also be ranked in the same manner.

8.6.3.1 Sensitivity and supportive analyses of primary efficacy endpoints

EU and RoW primary endpoint

- A sensitivity analysis will be performed by excluding patients whose baseline value is ≤70 µmol/L using the same approach as for the primary analysis.

- Logistic regression will be performed as a supportive analysis for this endpoint. The model will include treatment arm, baseline value, and stratification factors, i.e., PFIC type and age category. If the logistic model cannot converge, then the age category with small sample size may be pooled to handle the convergence of the modeling.

- A tipping point analysis will be conducted to evaluate the influence of missing data at the end of treatment. To start, one patient on placebo arm with a missing value at the end of treatment is assumed as a responder and the primary analysis will be re-run to see if the active treatment group is still significantly better compared to the placebo arm. This will be performed for each patient on placebo arm with a missing value at the end of treatment to see how the result changes, and the worst result will be reported for this step (i.e. the worst scenario). If the result is still significant then this process will continue by assuming any 2 patients on placebo arm with missing values as responders until the result becomes non-significant.

US primary endpoint

- An additional method is used to calculate the proportion of positive pruritus assessments up to the last report day before or on Day 168. In other words, the denominator to calculate the proportion will be (the last report day x 2). All missing data before or on the last report day will be considered as negative assessments. This estimated proportion of positive pruritus assessments at the patient level over the 24-week treatment period based on the Albireo ObsRO instrument will be analyzed using the same approach as for the primary analysis.

- To address the concern that results from subjects with a baseline AM or baseline PM average score less than 2 may be difficult to interpret, additional supportive analysis will exclude patients from the primary endpoint analysis who have a baseline AM or baseline PM average score that rounded to 0 or 1.

- A supportive analysis of pruritus assessments as a continuous endpoint will be conducted. The analysis will be performed by considering the impact of intermittently missing data as well as intercurrent events.

  The monthly (28-day) average in change from baseline for AM and PM, respectively will be calculated. There is a total of 6 monthly averages for AM assessments and 6 monthly averages for PM assessments over the 24-week treatment period. A mixed-effect model for repeated measures (MMRM) will be used to compare treatment effects at the last 28 days for AM and PM, respectively. Since missing not at random (MNAR) may occur, missing monthly average will be imputed with data from control arm completers using multiple imputation (MI). If there are subjects in the active arms who discontinued the study due to, e.g., relocation, their missing values may be imputed using active arm completers. To simplify the MI procedure, the description below only pertains to imputation using placebo completers. The MMRM model will include baseline score, treatment group, time (in month), treatment-by-baseline interaction, treatment-by-time interaction and stratification factors (i.e., PFIC type and age category). Unstructured covariance matrix will be used.

  The following procedures in MI will be followed:

  1. Intermediate missing values are imputed by the Markov Chain Monte Carlo (MCMC) imputation model, so only monotone missing values exist after imputation. 200 imputed data

sets will be created.

2. Subjects are classified to 3 subgroups: a. Completers in the control arm without a missing average at the last 4 weeks. b. Subjects with a missing average at the last 4 weeks in any arm, and c. Completers in the active arm without a missing average in the last 4 weeks. Subjects in subgroup 'a' will be used to impute missing values for subjects in subgroup 'b.'

3. The monotone missing values are imputed chronologically using SAS Proc MI. This is done by each imputed data set from step 1.

4. After the imputation, all subjects in the 3 subgroups are pooled and the data are sorted by the imputation ID from step 1.

5. MMRM is used to compare treatment effect in the last 4 weeks. This is done by the imputation IDs from step 1.

6. Proc MIAnalyze is used to summarize and conclude the results from step 5.

- If the comparison result from the analysis of MMRM with multiple imputation is significant, a tipping point analysis will be performed by reducing the imputed scores in placebo patients by 0.1 each time until the result is non-significant. In addition, the imputed scores in active patients will be increased by 0.1 each time till the result is non-significant.

- The MMRM with multiple imputation will be conducted for monthly average of change from baseline in AM and PM scores, and for monthly average of change from baseline in daily worst score.

8.6.4 Analyses of secondary efficacy endpoints

All secondary analyses will be based on the FAS. The analysis based on the PP populations will be performed for the 3 selected efficacy endpoints as specified below.

The change in the secondary endpoints such as serum bile acids, ALT, growth, Albireo PRO and ObsRO sleep parameters will be summarized by visit using descriptive statistics. Change in growth (weight for age and BMI for age) will also be displayed using graphical presentations. The summary of change in the secondary endpoints such as serum bile acids, ALT and growth will also be provided based on the PP analysis set.

Change in serum bile acids, change in ALT, change in growth will also be analyzed using a MMRM model. The model will include terms for baseline, PFIC type, age category, treatment, visit, treatment-by-baseline interaction and treatment-by-visit interaction. Unstructured covariance matrix will be used. The analysis of serum bile acids may be performed based on log transformed values of serum bile acids if deemed appropriate.

The analysis of growth data will be based on calculated values using the software or methods from the Centers for Disease Control and Prevention (CDC) website for patients with age ≥ 2 years old and from the WHO website for patients with age < 2 years old. CRF collected growth data will be listed.

The proportion of responders at Weeks 12 and 24 based on the ObsRO instrument will be analyzed using the same model as specified for the primary analysis for EU and RoW. The definition of the responder is a patient who reports a decrease in pruritus score from unrounded baseline equivalent to or greater than the threshold of meaningful change estimated from the blinded psychometric analysis; the results of the blinded psychometric analysis across all anchors support a threshold of a 1.0 point for AM, PM and AM and PM scratching scores. Section 8.9.1 includes more details about the blinded psychometric analysis. Specifically, the averaged pruritus score during weeks 11 and 12 and during weeks 23 and 24 will be used to calculate the proportion of patients achieving meaningful reduction at Week 12 and Week 24 from the baseline score. Similarly, the responder analysis will be conducted based on monthly scores at Week 12 and Week 24. Moreover, the analysis will be performed by AM, PM separately and AM & PM combined. For treatment completers whose last eDiary report day is before Week 24, if there is not enough eDiary data to calculate bi-weekly or monthly average scores at week 24, they will be excluded from analysis.

The proportion of individual assessments meeting the definition of a positive pruritus assessment at the subject level from Weeks 0-4, Weeks 0-8, Weeks 0-12, Weeks 0-18, Weeks 0-24, respectively, and the proportion of positive pruritus assessments at each 4-week interval over the 24-week treatment period for AM or PM, respectively will be analyzed using the same model as specified for the primary analysis for US. This analysis will be performed for PRO/ObsRO instruments separately.

The number and percent of patients achieving positive pruritus assessment for more than 50% of the time will be summarized by treatment group. Both AM and PM pruritus assessments will be included in the analysis of this endpoint. The proportion of patients achieving positive pruritus assessment for more than 50% of the time will be analyzed using the same model as specified for the primary analysis for EU and RoW. This analysis will also be performed separately for AM pruritus assessments only and PM pruritus assessments only.

A cumulative distribution function (CDF) plot showing proportions of patients achieving positive pruritus assessment for more than X% of the time, for X from 0 to 100%, by treatment group will be provided.

Number and percent of patients undergoing biliary diversion surgery and/or liver transplant will be summarized by using descriptive statistics. Kaplan-Meier curves will be used when appropriate for time to event data. Median event-free times and associated 95% confidence intervals will be calculated for each treatment group using Brookmeyer and Crowley methodology and a log-log transformation for constructing confidence intervals.

8.6.5 Analyses of exploratory endpoints

The exploratory efficacy variables are listed in Section 4.3. All the exploratory analyses will be based on the FAS.

Exploratory variables including Global Symptom Relief, biochemical markers and measures of bile acid synthesis, Albireo PRO and ObsRO itching/scratching severity scores, additional Albireo PRO and ObsRO sleep parameters, PELD/MELD score, APRI score, and the Fib-4 score will be analyzed descriptively for categorical and continuous data. For continuous data, the change from baseline will be analyzed in addition to the presentation of actual visit values. For Albireo PRO and ObsRO itching/scratching severity scores, unrounded baseline will be used to calculate change from baseline. For categorical data, shift tables or frequency and percentages of patients will be presented as appropriate. A line graph of itching/scratching daily severity scores of Albireo PRO and ObsRO over time for each patient will be provided.

For continuous data, such as PELD/MELD score, APRI score, and Fib-4 score, biochemical markers and measures of bile acid synthesis, the change from baseline to Week 24 will be analyzed descriptively in addition to the actual visit values. Global Symptom Relief data at Weeks 4, 12, and 24 will be summarized as categorical variables. Time points (visit window) are specified in Appendix B.

Comparisons of the change from baseline at Week 24 in the PedsQL total score (calculated as average score of all answered items) between the treatment groups will be conducted using analysis of covariance (ANCOVA). The model will include terms for baseline, PFIC type, age category, and treatment. The analysis will be conducted based on the total scores reported by child (>=5 years of age) and by parent (only including patients >=5 years of age) separately. If the sample size based on child reported data (>=5 years of age) is less than 10, then ANCOVA will not be conducted. The PedsQL total score of the family impact module will be analyzed similarly. The total score and the domain scores will be summarized descriptively.

Change from baseline to Week 24 in stage of liver fibrosis measured by Fibroscan® will be analyzed descriptively. The frequencies of stages may be summarized by shift tables. Steatosis grade analysis will be performed in a similar fashion.

8.6.6 Analyses based on exit survey data

The exit survey in Appendix F contains 3 questions and will be done by patients and caregivers separately. The answer Yes in Questions 1 and 2 will be verified with Question 3 for a positive and meaningful change.

The exit survey has the following questions:

1. Have you or your child experienced change from the study drug at the end of the study? Yes/No;

2. Was the change meaningful? Yes/No;

3. In what way was the change meaningful?

The exit survey was implemented in the middle of the trial so data from the exit survey will not be available for all patients. Analyses on exit survey are listed below. Analyses will be repeated considering missing data as No to Questions 1 and 2, and considering missing data as Yes to Questions 1 and 2.

The number and percent of patients with the following categories will be provided by treatment and overall:

- Q1 answer=Yes and Q2 answer=Yes
- Q1 answer=No and Q2 answer=No
- Q1 answer=Yes and Q2 answer=No

The exit surveys will be answered by both patients (≥8 yrs) and caregivers. Exit survey data will be provided in a data listing.

The following analyses on the exit survey will be performed using the descriptive summary based on pruritus assessments of Albireo ObsRO instrument in the last 2 weeks while patients are still on the study medications.

- For patients who answered Yes to Questions 1 and 2 in the exit survey, changes from baseline in the average scratching score based on the Albireo ObsRO instrument in the last 2 weeks before treatment termination

- For patients who answered No to Questions 1 or 2 in the exit survey, changes from baseline in the average scratching score based on the Albireo ObsRO instrument in the last 2 weeks before treatment termination

The summary based on the pruritus assessments of Albireo PRO instrument will also be provided.

8.7 Safety analyses

All definitions relative to safety endpoints are detailed in Section 4.4.

All safety analyses will be based on the Safety Analyses Set (as treated patients) and will be performed for all safety variables specified below and summarized by treatment group (odevixibat 40 µg/kg/day, odevixibat 120 µg/kg/day, odevixibat 2 doses combined, and placebo).

For each safety variable, the last value collected prior to first dose of study drug will be used as the baseline for all analyses.

No inferential statistical analysis of safety endpoints will be performed.

8.7.1 Adverse events

All adverse events (AEs) will be classified by Primary System Organ Class (SOC) and Preferred Term (PT) according to MedDRA version 23.0 throughout the study.

AEs will be classified as treatment emergent adverse events (TEAEs), defined as follows:
An AE (classified by preferred term) that occurs during the Treatment Period will be considered a TEAE if it has a start date on or after the first dose date of study drug, or it has a start date before the date of the first dose date of study drug, but worsened in severity on or after the date of the first dose date of study drug.

AEs with missing start dates, but with stop dates either overlapping with the treatment period or missing, will be considered TEAEs. A TEAE with missing drug-relationship will be considered as related. A TEAE with missing severity will be considered as severe. Details for imputing missing or partial start dates of adverse events are described in Appendix D.

For counting of the number of AEs, if 2 AE records have the same preferred term, and the start date of the 2nd AE is the same as or next day after the end date of the 1st AE, then they will be counted as one AE only.

An overall summary of the incidence of TEAEs (number of patients with any events and number of events, if applicable) will include the following:

- All TEAEs
- Drug-related TEAEs (AE will be defined as drug-related if causality is either probable, possibly or definitely)
- Severe TEAEs
- TEAEs leading to study discontinuation
- Serious TEAEs
- Drug-related serious TEAEs
- TEAEs leading to death
- Any hepatic TEAE
    a. Any liver-related TEAE (events that, per the Investigator, are considered related to PFIC)
    b. Any suspected drug-induced liver injury (DILI) TEAEs (as adjudicated by the Data Safety Monitoring Board [DSMB] and defined in Section 8.7.1.2)
    c. Any TEAE of liver decompensation (as defined in Section 8.7.1.2)
    d. Any TEAE in the standardised MedDRA query (SMQ) of *Drug Related Hepatic Disorders; Severe Events Only*
- Any Fat-Soluble Vitamin Deficiency TEAEs refractory to clinically recommended vitamin supplementation

- Any Clinically Significant Diarrhea TEAEs

TEAEs (number of patients with any events and number of events, if applicable) by SOC and PT in each treatment group will be tabulated for the following:

- TEAEs by SOC and PT including fat-soluble vitamin deficiencies, diarrhea, hepatoxicity (as defined in Section 8.7.1.1). Note that hepatotoxicity will also be included with the tabulation of hepatic events
- All TEAEs by preferred term by descending incidence in Odevixibat All Doses column
- TEAEs leading to study discontinuation by SOC and PT
- Serious TEAE by SOC and PT
- Drug-related serious TEAE by SOC and PT
- TEAEs leading to death by SOC and PT
- Common TEAEs (≥10% in any treatment group)
- Hepatic TEAEs
    a. Liver-related TEAEs (events that, per the Investigator, are considered related to PFIC) by SOC and preferred term
    b. Suspected DILI TEAEs (as adjudicated by the DSMB and defined in Section 8.7.1.2)
    c. TEAEs of liver decompensation (as defined in Section 8.7.1.2) by SOC and preferred term
    d. Any TEAE in the standardised MedDRA query (SMQ) of *Drug Related Hepatic Disorders; Severe Events Only*

Summary tables for number of patients with any TEAEs by SOC and PT by severity (mild, moderate, severe) and by causality (related [possibly, probably and definitely] vs unrelated [unlikely and unrelated]), will also be provided during the treatment period. AEs with the worst severity will be used in the by-severity summaries. Similarly, AEs with the worst causality (most related to treatment) will be used in the by-causality summaries. If severity or causality is missing, data will be imputed to the worst category.

When a patient has the same adverse event, based on preferred terminology, reported multiple times in the treatment period, the patient will only be counted once at the preferred terminology level in adverse event frequency tables. Where a patient has multiple adverse events within the same SOC in the treatment period, the patient will only be counted once at the SOC level in adverse event frequency tables.

In the AE summaries, AEs will be sorted by alphabetically for SOC and PT. In addition, the numbers of patients with liver-related mortality, liver decompensation events, and all-cause mortality will be presented using descriptive statistics. Kaplan-Meier curves may be used for time to event data (time taken to experience liver decompensation event, all-cause mortality, and biliary diversion surgery in weeks).

All AEs (including pre-treatment and post-treatment AEs), SAEs and deaths will be listed. Separate listings for AEs leading to dose interruption, and AEs of interest (Fat-Soluble Vitamin Deficiency, Diarrhea, and Hepatic AEs as defined below) will also be provided.

8.7.1.1 Definition for TEAEs of Fat-Soluble Vitamin Deficiency, Diarrhoea, and Hepatotoxicity

The following TEAEs have been defined based on the population under study:

- New or worsening of fat-soluble vitamin deficiency refractory to clinically recommended vitamin supplementation.

  All Investigator-reported verbatim terms related to decreases in vitamin levels or vitamin deficiency (e.g. preferred term of hypovitaminosis, Vitamin A decreased, Vitamin A deficiency) with the relevant concomitant medication records reviewed by the Medical Monitor, queried as needed, and the appropriate MedDRA preferred terms reported.

- Clinically significant diarrhoea, defined as any of the following:

  - Diarrhoea that persists for 21 or more days without any other aetiology based on medical review of other concurrent AEs for possible other causes of the diarrhoea or diagnostic testing (e.g. viral infections)

  - Reported by the Investigator as severe in intensity or reported as an SAE due to the requirement for hospitalisation or as an important medical event

  - Diarrhoea with concurrent dehydration requiring treatment with oral or intravenous rehydration and/or other treatment intervention based on medical review of AEs and concomitant medications

- Hepatoxicity: based on the SMQ of *Drug Related Hepatic Disorders; Severe Events Only*

For each of these categories, tabular summaries will be presented by MedDRA SOC and preferred term; the tables will also include the overall incidence of these AEs. Listings of patients with each of the events as defined will be provided.

8.7.1.2 Definitions and Determination of Hepatic Events

A DSMB was formed to independently assure the safety of patients enrolled in Study A4250 008, as well as to evaluate the integrity of study conduct and the data generated. The DSMB reviews the safety data at regular, pre-defined intervals and on an ad hoc basis as needed and makes recommendations regarding patient safety and study continuance (continuation, modification, or termination of the study). The DSMB is comprised of three paediatric hepatologists and an unblinded biostatistician.

As requested by the FDA, the DSMB has been chartered to review data from patients with hepatic AEs, including cases of suspected DILI and patients with liver decompensation events as defined below. The DSMB also reviews events that are reported in the SMQ of *Drug Related Hepatic Disorders; Severe Events Only*. The DSMB reviews each case and provides their expert opinion on the aetiology of the event.

For this review, Albireo prepares slides for presentation at the DSMB meeting on each patient who meets the criteria for hepatic event adjudication. Relevant liver-related laboratory values over time and a narrative summary of relevant information is provided to the DSMB. During the open session of the meeting, an Albireo physician reviews each case with the DSMB members, responds to questions, and/or obtains any additional information requested by the DSMB. Albireo's assessment of the hepatotoxicity aetiology is documented for each event. During the closed session, the DSMB independently assesses the event aetiology which is documented on a Hepatic Event Adjudication Form that is attached to the meeting minutes. If the DSMB requests follow-up information, this is provided at the next scheduled meeting. Treatment assignment remains blinded for this review. The Adjudication Process Document outlines the events that will be adjudicated.

8.7.1.2.1 Suspected Drug-Induced Liver Injury

As outlined in the protocols for Study A4250 008, patients with laboratory criteria that meet any of the following are considered suspected events of DILI and undergo review and adjudication of the event aetiology by the DSMB:

- ALT or AST ≥5 × upper limit of normal (ULN) if ALT or AST is normal at Baseline, or an absolute threshold of 800 U/L, whichever comes first
- ALT or AST ≥3 × Baseline if ALT or AST is abnormal at Baseline, or an absolute threshold of 800 U/L, whichever comes first
- ALT or AST ≥3 × ULN or 800 U/L, whichever comes first, and total bilirubin >2 × ULN
- Doubling of total bilirubin if total bilirubin was <3 mg/dL at Baseline
- Increase in total bilirubin by >3 mg/dL if total bilirubin was ≥3 mg/dL at Baseline
- INR increase >1.5 if INR was normal at Baseline and increase is refractory to Vitamin K administration
- INR increase by >0.4 if INR was abnormal at Baseline and increase is refractory to Vitamin K administration
- Any increase in total bilirubin and transaminases if accompanied by either a symptom of clinical hepatitis (vomiting, nausea, right upper quadrant pain) or immunological reaction (rash or 5% eosinophilia)

8.7.1.2.2 Liver Decompensation Adverse Events

Events either identified by Investigators or meeting the Albireo definition of liver decompensation will undergo review and adjudication by the DSMB. Patients who meet either of the following criteria undergo review and adjudication of the event aetiology by the DSMB for liver decompensation:

- INR elevation >1.5 that is refractory to vitamin K administration
- In a patient with portal hypertension and cirrhosis, transition to decompensated cirrhosis evidenced by any of the following:
    - Presence of ascites
    - Hepatorenal syndrome
    - Portopulmonary hypertension
    - Hepatopulmonary syndrome
    - Variceal haemorrhage
    - Hepatic encephalopathy

8.7.1.2.3 SMQ Drug-Related Hepatic Disorder AEs:

Events in the Standardised MedDRA query SMQ *Drug Related Hepatic Disorders; Severe Events Only (SMQ No. 20000007)* that are not captured in any of the above parameters will be presented to the DSMB for review.

The data of all adjudicated hepatic events provided by the DSMB will also be listed.

8.7.2 Clinical laboratory evaluations

Descriptive statistics for clinical laboratory values (in SI units for all tests, and in conventional units for selected tests (Appendix H)) and absolute changes from baseline at each post-baseline visit (Appendix B) will be presented. The change to the last visit will also be summarized. A shift table from baseline to the highest or lowest value for quantitative variables will be presented by treatment group for the clinical laboratory parameters listed in Table 2.

Central laboratory data will be used for the summary. If central laboratory data are not available due to COVID-19 or other reasons, local laboratory data will be used.

Figures for ALT will be prepared as follows:

- Mean change from baseline over time; all the treatment groups; one line per treatment group.
- Individual change from baseline over time; treatment group 40 µg/kg/day odevixibat; one line per patient.
- Individual change from baseline over time; treatment group 120 µg/kg/day odevixibat; one line per patient.
- Individual change from baseline over time; treatment group placebo; one line per patient.

The clinical laboratory test results will be listed.

Table 2: Laboratory Parameters

Clinical Chemistry	Hematology	Urinalysis	Other Labs
• Albumin • ALT • Alkaline phosphatase (ALP) • AST • Bilirubin – total and direct • Calcium • Chloride • Creatinine • Creatine kinase • Gamma-glutamyl transferase (GGT) • Potassium • Sodium	• Hematocrit • Hemoglobin • Platelet count • Red blood cell count • White blood cell count and differential (neutrophils, eosinophils, basophils, lymphocytes, and monocytes)	• Blood • Glucose • Ketones • Leukocytes • Nitrites • pH • Protein	• Vitamin A and E • 25-Hydroxy vitamin D • International normalized ratio (INR) • Alfa-fetoprotein (AFP) • Prothrombin Time

Furthermore, a listing of patients meeting the following criteria for liver monitoring will be provided:

- (ALT or AST ≥ 3x Baseline or >= 800 U/L) and TB > 2 ULN at the same visit
- ALT or AST >= 10 ULN or >= 5x Baseline or >= 800 U/L in presence of normal LDH and CPK
- ALT or AST >= 5 ULN (Normal Baseline) or >= 3x Baseline (Abnormal Baseline) or >= 800 U/L
- INR > 1.5 (Normal Baseline) or increased by > 0.4 relative to Baseline (Abnormal Baseline)
- Total bilirubin (TB) >= 2x Baseline (Baseline TB < 3 mg/dL) or increased by >= 3 mg/dL relative to Baseline (Baseline TB >= 3 mg/dL)

8.7.3 Vital signs

Descriptive statistics for vital signs (temperature, systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, weight, height and BMI) and their changes from baseline at each post-baseline visit will be presented by treatment group. The change to the last visit will also be summarized. A shift table for the number of patients with changes from baseline to the highest or lowest value (Appendix G) will be provided by treatment group during the treatment period.

Vital signs data will be listed.

8.7.4 Physical examinations

All physical examination data (general appearance, eyes, ears, nose, throat, head/neck/thyroid, lymph nodes, cardiovascular, lungs/chest, abdomen, genitourinary, extremities, skin, musculoskeletal, neurological, and other body systems) will be summarized at each visit with percentages and frequencies along with abnormalities by treatment group.

Abdominal ultrasound assessments for the liver, including Riedel's lobe, the portal vein and the spleen will be also summarized descriptively for each visit.

Skin assessments on a 5-point scale (0-4 from no evidence of scratching to cutaneous bleeding, haemorrhage, scabbing) for face, right arm, left arm, right leg, left leg and torso will be summarized at each visit with percentages and frequencies by treatment group.

Physical examination and skin assessment data will be listed (including pre-treatment and post-treatment results).

8.7.5 Pharmacokinetics (PK)

PK samples will only be drawn at visits 4 and 9 (only for children with body weight >10 kg) and if an SAE occurs, as close to the onset of an SAE as possible and if volume limitations permit. Samples will be processed and transported to a laboratory per instructions in the laboratory manual. Plasma concentrations will be listed. No PK parameters will be derived.

8.8 Subgroup analyses

Subgroup analyses will be performed for each of the 3 age groups (6 months to 5 years, 6 to 12 years, and 13 to 18 years), PFIC 1 and 2, region (US, EU and RoW), sex, race, ethnicity, baseline serum bile acid (≥250 and <250 µmol/L), Child-Pugh classification, Bile Salt Export Pump (BSEP) type of PFIC 2 patients, and the use of UDCA, rifampicin (alone or either one). Subgroup analyses may be conducted for hepatic impairment classification per NCI Organ Dysfunction Working Group (ODWG) if appropriate. Statistical analysis will be performed only when the sample size is ≥10 in each treatment group. If the sample size is <10 in either treatment group, only summary statistics will be provided; the p-value will not be reported. Due to small sample size in subgroups, all subgroup analysis will not be stratified by the stratification factors.

Analyses will be provided for the following parameters:
- The proportion of positive pruritus assessments at the subject level over the 24-week treatment period (primary endpoint)
- s-BA responder (primary endpoint)
- Change from baseline in s-BA, ALT, growth (secondary endpoints)

8.9 Blinded analysis

8.9.1 Pruritus Responder Definitions

Blinded analyses of Albireo ObsRO and PRO eDiary data were to be performed after 50% or more of the planned patients have completed the Week 24 Treatment Period. The actual number of 24-week treatment completers included in the analyses was 45 (~73%). The blinded analyses were used to estimate a threshold of clinically meaningful change (i.e., the responder definition) in Albireo ObsRO and PRO pruritus scores. Please refer to the psychometric analysis (Pruritus Evidence Dossier) for details. The analyses were performed by a group [redacted], independent from both the study team and the sponsor.

8.9.2 Data Safety Monitoring Board

The DSMB will receive data in the form of tables, figures and listings (provided by unblinded DSMB statistician from [redacted] who is independent to the [redacted] study team). The requirement for blinded or unblinded data is defined in the DSMB charter. The data provided to the DSMB will include, but is not limited to, demographics, baseline characteristics, medical and surgical history, prior and concomitant medications, AE and SAE data (by SOC and PT and by maximum intensity), laboratory data, liver monitoring, vital signs, abdominal ultrasound and patient disposition data.

9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

The changes to the planned analysis in the protocol are listed below.

1. Endpoints

(a) Added one secondary efficacy endpoint.

- Number of patients achieving positive pruritus assessment for more than 50% of the time during 24-week treatment period

(b) Modified one secondary efficacy endpoint

Protocol: Proportion of responders for pruritus scores at Weeks 12 and 24 based on the Albireo PRO and ObsRO instruments.

SAP: Proportion of responders for pruritus scores at Weeks 12 and 24 based on the ObsRO instrument.

(c) Added the following exploratory efficacy endpoints.

- For patients who answered Yes in Questions 1 and 2 in the exit survey, changes from baseline in the average scratching score based on the Albireo ObsRO instrument in the last 2 weeks before treatment termination

- For patients who answered No in Questions 1 or 2 in the exit survey, changes from baseline in the average scratching score based on the Albireo ObsRO instrument in the last 2 weeks before treatment termination

(d) Removed the following exploratory efficacy endpoints.

- Liver-related mortality and liver decompensation events
- All-cause mortality

(e) Added the following safety endpoints.

- Liver-related mortality and liver decompensation events
- All-cause mortality

2. Efficacy Analyses

(a) Changed the definition of fasting s-BA baseline
Protocol: Fasting s-BA baseline will be calculated as the average of the value at randomization and the previous visit (Clinic Visit 2).

SAP: Fasting s-BA baseline will be calculated as the average of the last 2 values prior to the first dose. The reason to change is that some patients may have unscheduled assessments between Clinic Visit 2 and Randomization Visit.

(b) Added the details of strata pooling strategy for the CMH test

(c) Added rounded baseline score which will be used for the analysis of pruritus assessments

(d) Added analyses for the number of patients achieving positive pruritus assessment for more than 50% of the time during 24-week treatment period

(f) Added the following analyses for the US primary endpoint of pruritus assessments

- An additional method to calculate the proportion of positive pruritus assessments up to the last report day before or on Day 168. In other words, the denominator to calculate the proportion will be (the last report day x 2). All missing data before or on the last report day will be considered as negative assessments.
- Additional supportive analysis will exclude patients from the primary endpoint analysis who have a baseline AM or baseline PM average score that rounded to 0 or 1.
- Multiple imputation and tipping point analysis for pruritus assessments as a continuous endpoint.

(g) Added the following analyses for the EU and RoW primary endpoint of fasting s-BA responder endpoint

- Tipping point analysis to evaluate the influence of missing data at the end of treatment

(h) Added the following analysis for secondary endpoints

- Change in serum bile acids, change in ALT, change in growth will be analyzed using a MMRM model

(i) Changes in Subgroup Analysis

- Removed a subgroup: for patients who were not included in the previous Study A4250-003
- Added subgroups by region (US, EU and RoW), sex, race, ethnicity, baseline serum bile acid (≥250 and <250 µmol/L), Child-Pugh classification, Bile Salt Export Pump (BSEP) type of PFIC 2 patients, and the use of UDCA, rifampicin (alone or either one). Subgroup analyses may be conducted for hepatic impairment classification per NCI Organ Dysfunction Working Group (ODWG) if appropriate.

(j) Corrected Formula of PELD Score and APRI Score

Protocol: PELD Score = 4.80 * ln(total bilirubin) + 18.57 * ln(INR) - 6.87 * ln(albumin) + 4.36 (if patient <1 year: scores for patients listed for liver transplantation before the patient's first birthday continue to include the value assigned for age (<1 year) until the patient reaches the age of 24 months) + 6.67 (if the patient has growth failure [**≤2 SD**])

SAP: PELD Score = 4.80 * ln(total bilirubin) + 18.57 * ln(INR) - 6.87 * ln(albumin) + 4.36 (if

patient <1 year: scores for patients listed for liver transplantation before the patient's first birthday continue to include the value assigned for age (<1 year) until the patient reaches the age of 24 months) + 6.67 (if the patient has growth failure [**< -2SD**])

Protocol: APRI = [(AST in U/L) / (AST ULN in U/L)] / (Platelets in 10^9/L)

SAP: APRI = [(AST in U/L) / (AST ULN in U/L)] × 100/ (Platelets in 10^9/L)

(k) Changed the definition of negative pruritus assessments

Protocol: All planned assessments after the intercurrent events (premature treatment discontinuation, death or initiation of rescue treatments such as biliary diversion surgery or liver transplantation) will be counted as negative pruritus assessments.

SAP: All missing planned assessments after premature treatment discontinuation will be counted as negative pruritus assessments. All planned assessments after death or initiation of rescue treatments such as biliary diversion surgery or liver transplantation will be counted as negative pruritus assessments. Both on-treatment and off-treatment pruritus assessments up to Day 168 will be included for analysis. 50% rule will be applied to the calculation. Please see details in Appendix A.

10 REFERENCES

1. ICH Topic E3: Structure and Content of Clinical Study Reports (CPMP/ICH/137/95-adopted December 1995).
2. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).
3. https://www.unos.org/wp-content/uploads/unos/MELD_PELD_Calculator_Documentation.pdf; accessed on 25-Jul-2018
4. Memorial Sloan-Kettering Cancer Center (Understanding your Fibroscan Results) https://www.mskcc.org/cancer-care/patient-education/understanding-your-fibroscan-results
5. Washington University, Louisville Kentucky (Jewish Hospital Fibroscan Interpretation Assessment SOP). https://louisville.edu/medicine/departments/medicine/divisions/gimedicine/physician-resources/calculators-and-tools-files/fibroscan-interpretation-sheet
6. Kane JM, Skuban A, Ouyang J, et al. A multicenter, randomized, double-blind, controlled phase 3 trial of fixed-dose brexpiprazole for the treatment of adults with acute schizophrenia, Schizophr Res. 2015: 164(1-3):127-35.
7. Lehmacher W, Wassmer G, Reitmeir P. Procedures for Two-Sample Comparisons with Multiple Endpoints Controlling the Experimentwise Error Rate. Biometrics 1991:47, 511-521.
8. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. Biometrika 1976:63:655-660.
9. Kadane JB. Sums of Possibly Associated Bernoulli Variables: The Conway-Maxwell-Binomial Distribution. Bayesian Analysis. 2016; 11(2):403-420.
10. Aaron S. Mansfield, Michelle A. Rudek, Diana Vulih, Gary L. Smith, Pamela Jo Harris, and S. Percy Ivy (2016). “The effect of hepatic impairment on outcomes in phase 1 clinical trials in cancer subjects”.
11. Vollset, S. E., Hirji, K. F., and Elashoff, R. M. (1991). “Fast Computation of Exact Confidence Limits for the Common Odds Ratio in a Series of Tables.” Journal of the American Statistical Association 86:404–409
12. Peds cases - Pediatrics for Medical Students - Developed by Chris Novak and Peter Gill for PedsCases.com April 21, 2016 (http://www.pedscases.com/pediatric-vital-signsreference-Chart)

11 APPENDICES

Appendix A: Derived Variables

The table below provides the list of derived variables for demographic and baseline characteristics, various duration derivations, drug compliance, baseline derivations and other important derivations applicable for this study.

Variables	Formula
Demographic and Baseline characteristics	
Age (in years)	For patients not from France and Germany, age will be calculated based on date of birth. For patients from France and Germany, only birth year was collected, and July 1 was imputed in the eDC system. The CRF collected age months and age years are based on the imputed date of birth. For analysis purpose, age will be calculated based on collected age months and age years on CRF or the external file (primary source).
Body mass index (BMI) (kg/ m^2)	weight (kg)/[height (m^2)
PFIC diagnosis (in years)	Will be derived by SAS YRDIF function yrdif (date of diagnosis of PFIC, Date of informed consent, 'ACT/ACT');
Derivation of Duration	
Study day at any visit	Date of interest – date of first dose of study drug. One day is added if this difference is ≥ 0
Extent of Exposure (days)	Date of last study drug intake – Date of first study drug intake + 1
Overall Study Drug Compliance	
Compliance based on CRF data	100 × [(total number of capsules dispensed – total number of capsules returned) /(total number of capsules planned to be taken)]
Compliance based on eDiary data	100 x (total number of capsules taken/total number of capsules planned to be taken). This will be considered as the primary compliance rate. The number of capsules planned to be taken will be estimated based on patient's body weight at Visit 3 and Visit 6 per Table 2 in the protocol. The number of capsules planned to be taken before Visit 6 will be the same as the one estimated at Visit 3, and after Visit 6 will be the same as the one estimated at Visit 6.
Derivations for Efficacy Parameters	
Baseline (general)	The baseline value of a parameter is defined as the last assessment of that parameter prior to the first intake of study treatment, except as indicated below.

Baseline for proportion of patients experiencing at least a 70% reduction in fasting s-BA and for s-BA change from baseline	Fasting s-BA baseline will be calculated as the average of the last 2 values prior to the first dose (see also Appendix B). If only one non-missing value is available, it will be used as baseline.
End value for proportion of patients experiencing at least a 70% reduction in fasting s-BA and for s-BA change from baseline	The end value is the average of the values at Weeks 22 and 24 after the start of double-blind treatment. (Appendix B). If one value is missing, then the non-missing value will be used as the end value. If both values are missing, then the end value is missing.
Baseline, daily, weekly and monthly score by AM and PM respectively (scratching, itch severity)	For both the Albireo ObsRO scratching item and the Albireo PRO itch severity score: All non-missing AM scores from the period of 14 days prior to or on the first dose day of study medications will be averaged as baseline. All non-missing PM scores from the period of 14 days prior to the first dose day of study medications will be averaged as baseline. Baseline score will be considered missing if ≥8 out of 14 assessments in the 14 days are missing (i.e. 50% rule is applied based on the number of planned assessments). A weekly AM (PM) score will be calculated by averaging all non-missing AM (PM) scores in a week. A weekly score will be considered missing if ≥4 out of 7 AM (PM) scores in a week are missing. A bi-weekly AM (PM) score will be calculated by averaging all non-missing AM (PM) scores in 2 weeks. A bi-weekly score will be considered missing if ≥8 out of 14 AM (PM) scores in 2 weeks are missing. A monthly AM (PM) score will be calculated by averaging all non-missing AM (PM) scores in a month (28 days). A monthly score will be considered missing if ≥15 out of 28 AM (PM) scores a week are missing.
AM & PM baseline, daily, weekly and monthly score (scratching, itch severity)	For both the Albireo ObsRO scratching item and the Albireo PRO itch severity score: A daily AM & PM score will be averaged from the 2 ratings for each day. A daily score will be considered missing if both assessments are missing. A weekly score will be calculated by averaging all non-missing AM and PM scores in a week. A weekly score will be considered missing if ≥8 out of 14 assessments in a week are missing. A bi-weekly score will be calculated by averaging all non-missing AM and PM scores in 2 weeks. A bi-weekly score will be considered missing if ≥15 out of 28 assessments in 2 weeks are missing. A monthly score will be calculated by averaging all non-missing AM and PM scores in a month (28 days). A monthly score will be considered missing if ≥29 out of 56 assessments in a month are missing.

	All non-missing AM scores from the period of 14 days prior to or on the first dose day of study medications, and all non-missing PM scores from the period of 14 days prior to the first dose day of study medications will be averaged as the AM & PM baseline score. Baseline score will be considered missing if ≥15 out of 28 assessments in the 14 days are missing. One thing of note: Inclusion criteria #4: Patient must have history of significant pruritus and a caregiver-reported observed scratching in the eDiary average of ≥2 (on 0 to 4 scale) in the 2 weeks prior to randomization. The pruritus score for purposes of establishing eligibility was calculated by taking the average of the worse of the two scores for each day. The values for week one were averaged and the values for week 2 were averaged. The two weekly averages were then averaged to provide a single score. If this score was ≥2 the pruritus eligibility requirement was met. The calculation of pruritus eligibility score is different from the calculation of the baseline pruritus score defined in the SAP.
Average of change from baseline from each AM and PM score (scratching, itch severity)	The calculation below is based on the change from baseline in each AM and PM score. The values based on this calculation will be used for data analysis. Change from baseline (daily) will be calculated by averaging all non-missing values of change from baseline (AM and PM) in a day. A daily change from baseline will be considered missing if both AM and PM change from baseline in a day are missing. Change from baseline (weekly) will be calculated by averaging all non-missing values of change from baseline (AM and PM) in a week. A weekly change from baseline will be considered missing if ≥8 out of 14 change from baseline values in a week are missing. Change from baseline (bi-weekly) will be calculated by averaging all non-missing values of change from baseline (AM and PM) in a week. A weekly change from baseline will be considered missing if ≥15 out of 28 change from baseline values in a week are missing. Change from baseline (monthly) will be calculated by averaging all non-missing values of change from baseline (AM and PM) in a month (28 days). A monthly change from baseline will be considered missing if ≥29 out of 56 assessments in a month are missing.
Baseline, daily, weekly and monthly (sleep parameters, such as difficulty of falling asleep and staying asleep, tiredness, the number of awakenings)	For patient- and observer reported outcome scores of sleep parameters, the same approach above will be used as for scratching/itch severity by AM and PM respectively since there is just 1 rating per day.
Proportion of positive pruritus assessments	In the primary analysis of the pruritus endpoint, all intermittently missing assessments will be classified as negative pruritus assessments. Similarly, all missing planned assessments after discontinuation of study drug will be counted as negative pruritus assessments. All planned assessments after death or initiation of rescue treatments such as biliary diversion surgery or liver transplantation will be counted as negative pruritus assessments. Both on-treatment and off-treatment pruritus assessments up to Day 168 will be included for analysis.

	Rounded baseline score will be used for the primary analysis. Additional analysis will be conducted based on unrounded baseline score. In general, all 336 assessments planned over the 24-week treatment period will be included in the denominator of the endpoint definition. However, if a completer's last pruritus assessment is reported prior to Day 168, then (the last report day x 2) will be used as the denominator. The last report day will be determined for ObsRO and PRO instruments separately. Regarding the proportion of positive pruritus assessments for each 4-week interval, the proportion is calculated up to the last interval with at least 50% of planned assessments.
Derivations for Safety Parameters	
Adverse Events (AE) duration (days)	AE end date – AE start date +1
Treatment Emergent Adverse Event (TEAE)	An AE (classified by preferred term) that occurs during the Treatment Period will be considered a TEAE if it has a start date on or after the first dose date of study drug, or it has a start date before the date of the first dose date of study drug, but worsened in severity on or after the date of the first dose date of study drug.

Appendix B: Visit Window

The observation closest to the target day is the measurement used in the analysis for each visit. The following 3 tables of analysis visit windows will apply for laboratory parameters, questionnaires, vital signs, physical measurements and other efficacy parameters in the study, if not otherwise specified.

Table 3.1: Analysis Visit Windows (General)

Timing of assessment (days relative to Treatment)	Visit Name to display for Analysis	Target day	Study Day (Relative Day)
Screening (Days -56 to -1)	Baseline		<= -1
Day 1 Week 1	Baseline		1 (pre-dose)
Week 4 (± 3 days)	Week 4	28	Post-baseline - 42
Week 8 (± 3 days)	Week 8	56	43 - 70
Week 12 (± 3 days)	Week 12	84	71 - 105
Week 18 (± 3 days)	Week 18	126	106 - 140
Week 22 (± 3 days)	Week 22	154	141 - 161
Week 24/EOT (± 3 days)	Week 24	168	162 - (last dose day + 14)
4 Weeks Post Last Dose of Study Drug (± 3 days)	Follow-up	28 post last dose	≥Last dose day + 15

Table 3.2: Analysis Visit Windows for GIC/GIS, Physical Measurements and Selected Lab Tests

Timing of assessment (days relative to Treatment)	Visit Name to display for Analysis	Target day	Study Day (Relative Day)
Screening (Days -56 to -1)	Baseline		<= -1
Day 1 Week 1	Baseline		1 (pre-dose)
Week 4 (± 3 days)	Week 4	28	Post-baseline - 56
Week 12 (± 3 days)	Week 12	84	57 - 126
Week 24/EOT (± 3 days)	Week 24	168	127 - (last dose day + 14)
4 Weeks Post Last Dose of Study Drug (± 3 days)	Follow-up	28 post last dose	≥Last dose day + 15

Note: Physical measurements include physical examination, voluntary photography and skin examination. Selected lab tests include urinalysis, autotaxin, p-C4, Vitamins A & E and 25-hydroxy vitamin D.

Table 3.3: Analysis Visit Windows for PedsQL, Fibroscan, Abdominal Ultrasound and AFP

Timing of assessment (days relative to Treatment)	Visit Name to display for Analysis	Target day	Study Day (Relative Day)
Screening (Days -56 to -1)	Baseline		<= -1
Day 1 Week 1	Baseline		1 (pre-dose)
Week 24/EOT (± 3 days)	Week 24	168	71 - (last dose day + 14)
4 Weeks Post Last Dose of Study Drug (± 3 days)	Follow-up	28 post last dose	≥Last dose day + 15

For laboratory and non-laboratory parameters, if a patient has more than one measurement included within a window, the assessment closest to the target day will be used. In case of ties between observations located on different sides of the target day, the earlier assessment will be used. For laboratory parameters, central laboratory results will be used over local laboratory results if both are in a same analysis window.

Derivations for pruritus and other itching, scratching, sleep parameters measured by the Albireo ObsRO and PRO instruments will be derived based on the following analysis window. For analysis purposes, diary entries will be assigned to a study day based on the recorded date regardless of recorded time. Please also refer to the derivations in the table of derived variables in Appendix A for details.

Analysis Window	Treatment Period	Start Study Day	End Study Day
Baseline	Week 0 (randomization)+	Day -14, PM assessment	Day 1, AM assessment
Week 4	Week 4	Day 1, PM assessment	Day 29, AM assessment
Week 8	Week 8	Day 29, PM assessment	Day 57, AM assessment
Week 12	Week 12	Day 57, PM assessment	Day 85, AM assessment
Week 16	Week 16	Day 85, PM assessment	Day 113, AM assessment
Week 20	Week 20	Day 113, PM assessment	Day 141, AM assessment
Week 24	Week 24 (EOT)	Day 141, PM assessment	Day 169, AM assessment
Follow-up	Week 28 (Follow-up)	Date of EOT PM assessment	Date of EOT + 28 days AM assessment

Appendix C: Definition of Region Variable

Country	Country code	Region variable
Australia	AUS	RoW
Belgium	BEL	EU
Canada	CAN	RoW
France	FRA	EU
Germany	DEU	EU
Israel	ISR	RoW
Italy	ITA	EU
Netherlands	NLD	EU
Poland	POL	EU
Saudi Arabia	SAU	RoW
Spain	ESP	EU
Sweden	SWE	EU
Turkey	TUR	RoW
United Kingdom	GBR	EU
United States	USA	US

Appendix D: Handling of Missing or Incomplete Dates

Global statement:
If the imputed date is prior to the date of birth, then impute the missing date as date of birth.

Imputation rules for missing or partial AE start date are defined below:

If only Day of AE start date is missing:
If the AE start year and month are the same as that for the first dose date, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start day as the day of first dose date;
- If AE end date is prior to first dose date, then impute the AE start day as 1.

If Day and Month of AE start date are missing:
If AE start year = first dose year, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start month and day as the month and day of first dose date;
- If AE end date is prior to first dose date, then impute the AE start month as January and the day as 1.

If Year of AE start date is missing:
If the year of AE start is missing or AE start date is completely missing, then query the site with no imputation. Also compare the full (or partial) AE end date to the first dose date. If the AE end date is before the first dose date, then the AE should be considered as a pre-treatment AE. Otherwise, the AE will be considered as TEAE.

Imputation rules for missing or partial non-study medication start/stop dates are defined below:

Missing or partial non-study medication start date:

- If only day is missing, use the first day of the month.
- If day and month are both missing, use the first day of the year.
- If day, month and year are all missing, use the date of the day before the first dose date.

Missing or partial non-study medication stop date:

- If only day is missing, use the last day of the month.
- If day and month are both missing, use the last day of the year.
- If day, month and year are all missing, assign ‘continuing’ status to stop date

Appendix E: SAS Codes

1. Cochran-Mantel-Haenszel test

```
Proc freq data=< >;
  tables pfic*ageclass*trt*responder/cmh commonriskdiff;
  exact comor;
  ods output cmh=cmh (where=(AltHypothesis="General Association"));
run;
```

2. ANCOVA model

```
Proc mixed data=< >;
  class trt(ref='Placebo') pfic ageclass;
  model proportion of positive assessment = trt baseline pfic ageclass / solution;
  lsmeans trt / cl diff alpha=0.05;
  lsmestimate trt 'total odevixibat dose' 0.5 0.5 0 ;
  lsmestimate trt 'avg odevixibat dose effect - placebo' 0.5 0.5 -1/ cl alpha=0.05;
run;
```

Appendix F: Exit Survey

Protocol: A4250-005 Exit Questionnaire

Site: _ _ _ _ _
Patient Number: _ _ _
Date: (dd-mmm-yyyy) ____ /____ /____

- The questionnaire should be administered at the End of Treatment (EOT) visit
- Please ask Caregiver and Patient (if applicable; ≥8 yrs) the questions and transcribe the responses directly on the form. The responses require data entry in the InForm CRF.
- Guidelines/Tips:
  - It is best to ask Patient the questions first, preferably without Caregiver in the room to reduce bias
  - As the questions could be difficult for young patients to understand, it is acceptable to use any terms/phrases to help them comprehend what is being asked

Please ask Caregiver the following questions:

1. Have you or your child experienced change from the study drug at the end of the study?
   = Yes = No

If answered "yes", please ask the following question

2. Was the change meaningful?
   = Yes = No

If answered "yes", please ask the following question

3. In what way was the change meaningful?

   _____

Please ask Subject (≥8 years of age) the following questions:

1. Have you experienced change from the study drug at the end of the study?
   = Yes = No

If answered "yes", please ask the following question

2. Was the change meaningful?
   = Yes = No

If answered "yes", please ask the following question

3. In what way was the change meaningful?

   _____

Exit Questionnaire Version 1.0 16 July 2019 Albireo Pharma

Appendix G: Normal Reference Ranges for Vital Signs [12]

Heart Rate by Age (beats/minute) reference

Age	Awake Rate
Infant (<1 y)	100-190
Toddler (1-2 y)	98-140
Preschool (3-5 y)	80-120
School-age (6-11 y)	75-118
Adolescent (12-15 y)	60-100
> 15 y	60-100

Normal Respiratory Rate by Age (breaths/minute) reference:

Age	Respiratory Rate
Infants (<1 y)	30-53
Toddler (1-2 y)	22-37
Preschool (3-5 y)	20-28
School-age (6-11 y)	18-25
Adolescent (12-15 y)	12-20
> 15 y	12-20

Normal Blood Pressure by Age (mm Hg) reference:

Age	Systolic Pressure	Diastolic Pressure
Infant (<1 y)	72-104	37-56
Toddler (1-2 y)	86-106	42-63
Preschooler (3-5 y)	89-112	46-72
School-age (6-9 y)	97-115	57-76
Preadolescent (10-11 y)	102-120	61-80
Adolescent (12-15 y)	110-131	64-83
> 15 y	90-120	50-80

Normal Temperature Range by Method:

Method	Temperature (°C)
Rectal	36.6-38
Ear	35.8-38
Oral	35.5-37.5
Axillary	36.5-37.5
Temporal/Core*	35.8-38

*In the eCRF core and temporal temperature measurements can be ticked as methods. Rectal & Tympanic (Ear) fall in this group. Additionally, temporal temperature measurements (using a temperature scanner on the forehead) – approximate closely to core temperature measurements and therefore can be classified in this category. For that reason, the normal temperature range for core and temporal temperature measurements are defined as described.

Appendix H: SI and US Conventional Units of Clinical Laboratory Values

SERUM CHEMISTRY	SI UNIT	CONVENTIONAL UNIT
Analyte		
Alpha Fetoprotein	IU/mL	ng/mL
Direct bilirubin	µmol/L	mg/dL
Calcium	mmol/L	mg/dL
Chloride	mmol/L	mEq/L
Creatinine	µmol/L	mg/dL
Potassium	mmol/L	mEq/L
Sodium	mmol/L	mEq/L
Serum bile acid	µmol/L	mg/dL
Total bilirubin	µmol/L	mg/dL
Haematology	**SI Unit**	**Conventional Unit**
Analyte		
Haematocrit	ratio (L/L)	%
Haemoglobin	g/L	g/dL
Red blood count (RBC)	x10^12/L	x10^6/µL
Platelet count	x10^9/L	x10^3/µL
White blood cell count	x10^9/L	x10^3/µL
Fat Soluble Vitamins	**SI Unit**	**Conventional Unit**
Vitamin A	µmol/L	ug/dL
Vitamin E	µmol/L	mg/L
Vitamin D (25-dihydroxy)	nmol/L	ng/mL
Vitamin K	nmol/L	ng/mL
Urinalysis	**SI Unit**	**Conventional Unit**
Analyte		
Glucose	mmol/L	mg/dL
Ketones	mmol/L	mg/dL
Protein	mmol/L	mg/dL