SPONSOR:

Genkyotex

PROTOCOL NUMBER:

GSN000300

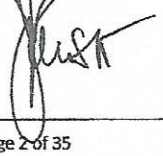
# STATISTICAL ANALYSIS PLAN

| | |
|---|---|
| Author: | Jean-Luc Beffy |
| Version: | Final version 2.0 |
| Date: | 15MAR2019 |

# 1 Cover and signature pages

| | |
|---|---|
| **Sponsor:** | Genkyotex |
| **Protocol Number:** | GSN000300 |
| **Study Title:** | A Double-Blind, Randomized, Placebo-Controlled Clinical Trial to Assess the Efficacy and Safety of Oral GKT137831 in Patients with Primary Biliary Cholangitis Receiving Ursodeoxycholic Acid and with Persistently Elevated Alkaline Phosphatase |
| **Document Version No** | Final version 2.0 |

**We, the undersigned, confirm that we have read, understood and agree to the content of this document and hereby authorise its approval.**

| **Name**<br>**Title, Organisation** | **Signature** | **Date** |
|---|---|---|
| Jean-Luc Beffy<br>Principal Biostatistician<br>Cmed CRS Ltd |  | 19 MAR 2019 |
| Dimoula Erakli<br>Senior Biostatistician<br>Cmed CRS Ltd |  | 19 MAR 2019 |
| Philippe Wiesel<br>Chief Medical Officer<br>Genkyotex Ltd |  | March 19 2019 |
| Yves Grumser<br>Safety Physician<br>Genkyotex SA |  | March 19 2019 |

# Table of Contents

# 2 List of Abbreviations and Definition of Terms

| | |
|---|---|
| AE | Adverse Event |
| AESI | Adverse Event of Special Interest |
| ALP | Alkaline Phosphatase |
| ALT | Alanine Aminotransferase |
| ANCOVA | Analysis of Covariance |
| APRI | AST to Platelet Ratio Index |
| AST | Aspartate Aminotransferase |
| BID | Twice Daily |
| CK-18 | Cytokeratin-18 |
| CSR | Clinical Study Report |
| DBP | Diastolic Blood Pressure |
| ECG | Electrocardiogram |
| eCRF | Electronic Case Report Form |
| ELF | Enhanced Liver Fibrosis |
| FDA | Food and Drug Administration |
| FIB-4 | Fibrosis-4 |
| GGT | Gamma Glutamyl Transferase |
| hsCRP | High Sensitivity C-Reactive Protein |
| ICH | International Conference on Harmonization |
| IgM | Immunoglobulin M |
| IMP | Investigational Medicinal Product |
| IPF | Idiopathic Pulmonary Fibrosis |
| ITT | Intent-To-Treat |
| IWRS | Interactive Web-Based Randomization System |
| LOCF | Last Observation Carried Forward |
| MSAP | Modeling and Simulation Analysis Plan |
| NADPH | Nicotinamide Adenine Dinucleotide Phosphate |
| OCA | Obeticholic Acid |
| OD | Once Daily |
| PBC | Primary Biliary Cholangitis |

| | |
|---|---|
| PD | Pharmacodynamic |
| PK | Pharmacokinetic |
| PP | Per Protocol |
| QTcF | Corrected QT Interval (Fredericia's Formula) |
| SAP | Statistical Analysis Plan |
| SBP | Systolic Blood Pressure |
| SMB | Safety Monitoring Board |
| TEAE | Treatment-Emergent Adverse Event |
| TSH | Thyroid-Stimulating Hormone |
| UDCA | Ursodeoxycholic Acid |
| ULN | Upper Limit Of Normal |
| VAS | Visual Analog Scale |
| WHO | World Health Organization |

# 3 Introduction

The purpose of this document is to describe the statistical methods, data derivations and data summaries to be employed in the analysis of study GSN000300 set up by Genkyotex in order to assess the efficacy and safety of oral GKT137831, a selective inhibitor of NOX 1 and 4 isoforms of the Nicotinamide Adenine Dinucleotide Phosphate (NADPH) oxidase family of enzymes, in patients with primary biliary cholangitis (PBC) receiving ursodeoxycholic acid (UDCA) and with persistently elevated alkaline phosphatase (ALP).
The preparation of this statistical analysis plan (SAP) has been based on International Conference on Harmonisation (ICH) E3 and E9 Guidelines and in reference to Protocol GSN000300 Version 4.0 (Amendment 3, 24 July 2018) and CRF Version 6.4 (12 February 2019).

# 4 Study Objectives

**PRIMARY OBJECTIVE**

The primary objective of the study is defined as follows:

- To evaluate the efficacy of oral GKT137831 in comparison with placebo, in subjects with PBC receiving UDCA and with persistently elevated serum ALP.

**SECONDARY OBJECTIVES**

The secondary objectives of the study are defined as follows:

- To evaluate the safety of oral GKT137831 in comparison with placebo, in subjects with PBC receiving UDCA and with persistently elevated serum ALP.
- To estimate the pharmacokinetics (PK) population of GKT137831 and explore any potential PK-PD (pharmacodynamics) relationships in this subject population.
- To explore any relationship between genetic parameters and therapeutic responses in a subset of subjects.

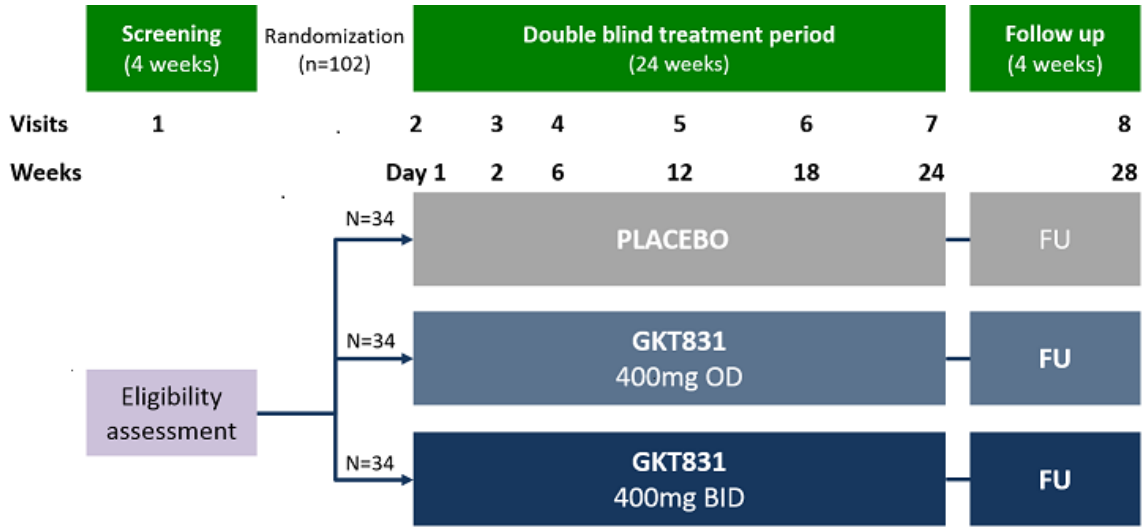
# 5 Study Design

## *5.1 STUDY DESIGN AND POPULATION*

The study population consists of approximately 100 subjects with PBC receiving a stable dose of UDCA and with persistently elevated ALP.
The study is a double-blind, randomized, placebo-controlled, multicenter, parallel group phase 2 trial assessing a 24-week period of treatment with oral GKT137831 administered in addition to standard of care medication (UDCA) in subjects with PBC. Subjects will be followed up for 28 days after the end of the treatment period.

An overview of the study design is presented in Figure 1 below.

Figure 1: Study Design Flowchart




### *5.2 STUDY TREATMENTS AND ASSESSMENTS*

Subjects will be assessed for their eligibility during the screening period (Visit 1), which will last up to 4 weeks, until the baseline/Day 1 visit (Visit 2).
Eligible subjects will be allocated to oral GKT137831 (400 mg Once Daily (OD) or 400 mg Twice Daily (BID) or placebo, according to a 1:1:1 randomization ratio stratified by disease severity level at study entry defined as baseline serum Gamma Glutamyl Transferase (GGT) < 2.5 x Upper Limit of Normal (ULN) or ≥ 2.5 x ULN.
Subjects will self-administer orally 400 mg OD or 400 mg BID of oral GKT137831 or matching placebo for a total of 24 weeks.
Baseline assessments will be performed at baseline/Day 1 (Visit 2). The 24-week treatment period will include assessments after 2 weeks of treatment (Visit 3), after 6 weeks of treatment (Visit 4), after 12 weeks of treatment (Visit 5), after 18 weeks of treatment (Visit 6) and after 24 weeks of treatment (End of Treatment/Visit 7). Subjects will be followed up to 28 days after the end of treatment (Week 28/Visit 8), totaling 6 post-baseline visits. Subjects who discontinue treatment before Week 24 visit will attend an Early Termination visit (premature end of treatment).
Pharmacokinetic samples will be taken at baseline, Week 2, Week 12 and Week 18 visits (see Protocol Version 3.0 - Section 8.8.2 for further details regarding the sampling time points).
Subjects will be taking a stable dose of UDCA at enrollment and will continue their UDCA treatment at a stable dose (no changes at all) during the treatment period.
A Safety Monitoring Board will oversee the conduct of the study to ensure the safety of

participating subjects (see Protocol Version 3.0 - Section 9.7 for more details). An interim analysis will be conducted when at least 80-90% of the planned number of subjects to be randomized in the study have completed their Week 6 visit (see Section 9). This IA analysis will assess demographics and efficacy data collected up to Week 6.

### *5.3 RANDOMIZATION AND BLINDING*

Approximately 100 subjects will be randomly allocated to placebo or one of the 2 active treatment arms, according to a 1:1:1 randomization ratio stratified at study entry by disease severity level defined as baseline serum GGT < 2.5 x ULN or ≥ 2.5 x ULN).
Accordingly, approximately 33-34 subjects will be allocated to each of the 3 treatment arms.
A master reproducible randomization list will be produced by or under the responsibility of Cmed accounting for block size and stratification. The Interactive Web-Based Randomization System (IWRS) will assign a unique randomization number in ascending, sequential order (with associated treatment arm) to the subject, based on the pre-determined randomization schedule. The system will assign the pre-determined blocks of randomization numbers for each stratification level to ensure similarly balanced treatment groups. Assignment will be in sequential order within blocks and stratification level. The investigator will enter the randomization number in the electronic case report form (eCRF).
The Sponsor, subjects, investigator staff, persons performing the assessments, data reviewers and statisticians will remain blinded to the identity of the study treatments.
The identity of the study treatments will be concealed by the use of Investigational Medicinal Products (IMPs) which are all identical in packaging, labeling, schedule of administration, appearance and odor.
Randomization data will be kept strictly confidential and be accessible only to authorized personnel (e.g., unblinded pharmacist or authorized designee), until unblinding of the trial as described in this SAP.
Unblinding will only occur for the following reasons:

- Subject emergencies (see Protocol Section 5.6.3 for more details)
- Scheduled and unscheduled safety reviews by the Safety Monitoring Board (SMB) (see Protocol Section 9.7 for more details)
- At the time of the interim analysis (see Section 9)
- At the end of the study.

### *5.4 SAMPLE SIZE JUSTIFICATION*

A sample size of 34 randomly allocated and treated subjects within each treatment group with an overall sample size of 102 subjects will have 80% power to detect a 28% difference in the means of the percent change from baseline in serum GGT. A Wilcoxon Mann-Whitney test has been used for the sample size estimate because this represents the worst-case scenario with regards to the statistical power. A standard deviation of 30 for the active GKT137831 group and a standard deviation of 40 for the placebo group have been assumed based on a recent phase 3 clinical trial of OCA [1].

An overall two-sided Type I error of 5% has been considered for the sample size calculation.
The Hochberg method will be used to adjust the alpha level for multiple comparisons. Therefore, the first dose level comparison at the final analysis will be tested against an alpha level of 0.04695 and the second dose level will be tested against an alpha level of 0.023475. The alpha level of 0.023475 has been used for the sample size calculation.
The sample size calculation assumes all subjects (including withdrawals) will be included in the analysis. As a result, there has been no adjustment for dropouts.

# 6 Statistical Considerations

The SAS system version 9.4 (or higher), will be used for all analysis, unless otherwise specified.

## *6.1 TIMING FOR THE FINAL ANALYSIS*

The final analysis will be performed according to a 2-stage process as detailed hereafter:

- **Stage 1: end of the double-blind treatment period**

The double-blind treatment period will be completed once the last subject last Week 24 visit will have been attended.
The date of the last subject last Week 24 visit will be referred to as the cut-off date for Stage 1.
All the data collected until the cut-off date will be taken into consideration for a first database lock.
As a result, all the efficacy data required for the efficacy endpoints defined in Section 8.7 will be included in this first locked database.
This first database lock is requested by Genkyotex to enable Cmed to produce and deliver Top Line tables (see the list in Appendix 1) as soon as all Week 24 visit efficacy data has been collected, cleaned, and locked. This first database lock will also include all Week 28 follow-up visit data which will have been collected until the cut-off date and cleaned prior to the database lock. The second database lock (see Stage 2 below) is to be carried out once all remaining follow-up (V8) data has been collected and cleaned. Week 28 follow-up visit data not included in the first database lock will be listed separately in the clinical study report and in any other relevant document.

- **Stage 2: end of the follow-up period**

The follow-up period will be completed once the last subject last Week 28 follow-up visit will have been attended. It will be considered as the end of the study.
The date of the last subject last Week 28 follow-up visit will be referred to as the cut-off date for Stage 2.
All the data collected until the cut-off date will be taken into consideration for a second database lock. It will be regarded as the final database lock.
Cmed will produce and deliver both the remaining planned efficacy outputs and all the other planned outputs to Genkyotex after the final database lock.

### *6.2 MISSING DATA HANDLING*

In the event of missing data at Week 24 visit for efficacy endpoints (see the list below), the Last Observation Carried Forward (LOCF) method will be applied by imputing Week 24 visit data with the last non-missing post-baseline data available at Week 12 visit (or at a later visit). If there is no non-missing post-baseline data or if the last non-missing post-baseline data is available at an earlier visit than Week 12 visit then the LOCF method will not be applied and no data will be imputed at Week 24 visit. Accordingly, the LOCF method will not be used for the interim analysis.
The above-described use of the LOCF method for the imputation of missing data at Week 24 visit is based on recently published studies that have shown that the reductions in liver function tests are close to maximal effects after 12 weeks of treatment [2].
The LOCF method will be applied for the following efficacy endpoints:

- Primary efficacy endpoint;
- Week 24-defined secondary efficacy endpoints (except those defined for ELF and liver stiffness);
- Week 24-defined tertiary efficacy endpoint for total bile acids;
- Week 24-defined exploratory efficacy endpoints (except those from the assessments of the metabolomics signatures and additional biomarkers of interest)

The complete case approach will be applied for the analysis of categorical variables using a worst-case imputation strategy.
For all other analyses and summaries, no imputation will be performed other than to complete partial dates using standard imputation techniques as described below.

### *6.3 PARTIAL DATE IMPUTATION*

The following rules should be used when modifying partial or missing dates for reporting purposes such as defining on treatment flags.
A permanent new date variable should be created if there is a requirement to be used in determining flags, sort orders and other derived variables needed for a table, listing or figure. Imputed date variable names will be defined in the derived dataset specifications.
Original (raw) date variables must not be overwritten. Imputed dates will not be displayed in the listings.

**General rules**

Adverse events (AE) or prior and concomitant medications are considered to have started at the earliest possible date and end at the latest possible date.

In case of partial dates with missing day:

For any AE or concomitant medication starting during the month of first dose, the partial start date will be imputed at the date of first dose, taking the worst case scenario.
For any AE or concomitant medication starting after the month of first dose, the start date will be imputed at the first day of the month.

For any AE or concomitant medication starting before the month of first dose, the start date will be imputed at the last day of the month**.**
For any AE or concomitant medication, a partial end date will be imputed at the last day of the month or at the date of study discontinuation, whichever occurs first.

In case of partial dates with missing day and month:

For any AE or concomitant medication starting during the year of first dose, the partial start date will be imputed at the date of first dose, taking the worst case scenario.
For any AE or concomitant medication starting after the year of first dose, the partial start date will be imputed at the first of January.
For any AE or concomitant medication started before the year of first dose, the partial start date will be imputed as the last day of the year.
For any AE or concomitant medication, a partial end date will be imputed at the last day of the year or at the date of study discontinuation, whichever occurs first

Some examples are given below (YYYY-MM-DD).
In most cases, start dates are imputed as first day of the month or first day of the year.

| Data Type | Start Date | Imputed Start Date | First dose | Date of disconti-nuation | End Date | Imputed End Date |
|---|---|---|---|---|---|---|
| Adverse Event, Prior/Concomitant Meds | 2017-02 | 2017-02-01 | 2016-12-17 | - | 2017-02 | 2017-02-29 |
| Adverse Event, Prior/Concomitant Meds | 2017-02 | 2017-02-03 | 2017-02-03 | - | 2017-02 | 2017-02-29 |
| Adverse Event, Prior/Concomitant Meds | 2017-02 | 2017-02-03 | 2017-02-03 | - | 2017-03 | 2017-03-31 |
| Adverse Event, Prior/Concomitant Meds | 2017-03 | 2017-03-01 | 2017-02-03 | - | 2017-03 | 2017-03-31 |
| Adverse Event, Prior/Concomitant Meds | 2017-01 | 2017-01-31 | 2017-03-03 | - | 2017-01 | 2017-01-31 |
| Adverse Event, Prior/Concomitant Meds | 2016 | 2016-12-31 | 2017-02-03 | 2017-03-16 | 2017-03 | 2017-03-16 |
| Adverse Event, Prior/Concomitant Meds | 2017 | 2017-02-03 | 2017-02-03 | 2017-04-08 | 2017-03 | 2017-03-31 |

Partial dates for initial diagnosis of PBC will be imputed as the 15$^{th}$ of the month if the month and the year are recorded, or the 1$^{st}$ of July if only the year is recorded.

Partial dates are not expected for efficacy-related assessments and death.
However, should a partial date be recorded for any efficacy-related assessment then this date will be imputed as the first day of the month if the month and the year are recorded or the latest between first of January and date of first dose if only the year is recorded.
In case of partial date for death, the date will be imputed as the day after the last visit/assessment date when the patient was known alive.

### *6.4 VISIT WINDOWING*

Visit windowing will be considered only for efficacy and pharmacodynamic assessments.
The halfway point corresponding to the visits will be considered in relation to the date of first dose. Efficacy assessments are planned to be conducted at Week 2, Week 6, Week 12, Week 18, Week 24 and Week 28 visits. The following windowing rules will be taken into account to establish the visit corresponding to each efficacy and pharmacodynamic assessment:

- Week 2: any efficacy or pharmacodynamic assessment performed between 2 to 28 days after the first dose;
- Week 6: any efficacy or pharmacodynamic assessment performed between 29 days to 63 days after the first dose;
- Week 12: any efficacy or pharmacodynamic assessment performed between 64 days to 105 days after the first dose;
- Week 18: any efficacy or pharmacodynamic assessment performed between 106 days to 147 days after the first dose;
- Week 24: any efficacy or pharmacodynamic assessment performed between 148 days to 182 days after the first dose;
- Week 28: any efficacy or pharmacodynamic assessment performed between 183 days to 203 days after the first dose;

If two (or more) assessments are recorded within a visit window, the closer (closest) to the halfway point will be taken into account for the analyses. If two assessments within a visit window are the same number of days from the halfway point, the second one will be taken into account for the summaries assuming that it is more reliable than the first one.
No programmatic windowing of visits will be considered for safety data. Post-baseline data will be presented according to the visit at which it was collected on the eCRF or as described in this SAP.

### *6.5 BASELINE*

Baseline is defined as the last non-missing value/result where assessment date is less than or equal to the date of first dose, unless otherwise specified for individual assessments.
Baseline will be determined based on all assessments, including additional assessments.
Change from baseline is defined as the difference between the post-baseline assessment value and the baseline value.

### *6.6 REPORTING GUIDELINES*

The following guidelines will be followed:

- **Page Orientation:** Landscape.
- **Post-text tables and listings**: will be generated in .lst and converted to rtf.
- **Post-text figures**: will be generated directly in .rtf.
- **Font**: Courier New font with minimum of 9 points font size.
- **Margins**: Left: 3.8 cm, Right: 2 cm, Top: 3 cm, Bottom 2 cm on A4 paper.
- **Columns headers** will be left aligned.
- **Treatment labels** will be the following and displayed in the following order, unless otherwise stated:
  - GKT831 400mg OD
  - GKT831 400mg BID
  - Placebo
  - Total
- **Visit labels**: the visit labels displayed in Table 1 will be used as required.

**Table 1: Visit Labels**

| Study Stage | CRF Visit | Tables, Figures and Listings Label |
|---|---|---|
| Screening | Visit 1 (Screening): Day -28 to -7 | SCR |
| Treatment phase | Visit 2 (Baseline): Day 1 | BAS |
| | Visit 3: Week 2 +/- 3 days | TRT W2 |
| | Visit 4: Week 6 +/- 3 days | TRT W6 |
| | Visit 5: Week 12 +/- 3 days | TRT W12 |
| | Visit 6: Week 18 +/- 3 days | TRT W18 |
| | Visit 7: Week 24 +/- 3 days | TRT W24 |
| Follow-up phase | Visit 8: Week 28 +/- 3 days | FU W28 |

- **Unscheduled visit / repeat assessments**: Data obtained at unscheduled or repeat assessments will be taken into account in baseline determination and efficacy summaries based on windowed visits. All other data from unscheduled or repeat assessment will not be included in summaries but only be presented in data listings, unless otherwise specified.
- **N:** The number of subjects in the specified population.
- **Treatment presentation:** The summaries will be presented by allocated or actual treatment group (see more details in Section 7) and overall for disposition, demographics and baseline characteristics, protocol deviations, concomitant medications, study drug exposure and compliance.
- **Continuous data** will be summarized using number of subjects (n), mean, standard deviation (SD), median, minimum value, maximum value and number of missing data (if there are any).
- **Categorical data** will be summarized using n and percentage based on number of non-missing data.
    - All categories will be presented, even if no subjects are counted in a particular category.
    - In case 1 or more subjects have missing data for the summary, the number of missing data will be presented as a separate category, labelled accordingly as 'Missing', if not otherwise stated.
    - Counts of zero in any category will be presented without percentage.
    - For AEs, medical history, prior and concomitant medications the counts are based on single counts of subjects with multiple events/treatments under same category, while the percentages are calculated using N.
- **Precision of summary statistics:**
    - Integer – Sample size (n, N) and number of missing data (if displayed);
    - One additional decimal place than reported/collected – mean, median, other percentile, confidence interval;
    - Two additional decimal places than reported/collected – standard deviation;
    - Same number of decimal places as reported/collected – minimum, maximum;
    - Percentages – one decimal place.
- **Study day** for an assessment will be calculated with reference to the date of first dose as Day 1.

  This will be calculated as follows:
  (assessment date - date of first dose) + 1 if the assessment date is on or after the date of first dose,
  (assessment date - date of first dose) if the assessment date is prior to the date of first dose.

- Data will be presented in listings by allocated treatment group. The order will be subject ID, visit, assessment date/time and assessment type/parameters (in the order collected on e-CRF, unless otherwise specified). For clinical laboratory results, the listings will be presented in the order of allocated treatment group, subject ID, parameter, assessment date/time, visit.
- Dates will be presented in format DD-MMM-YYYY.
- Version 4.03 of the NCI-CTC grading criteria (CTCAE v4) will be used for relevant tables [3].
- Latest version of the Medical Dictionary for Regulatory Activities (MedDRA) will be used for the adverse events coding. The version will be specified in the footnote of the corresponding TFLs.
- Latest version of the WHO-DRUG dictionary will be used for the medications coding. The version will be specified in the footnote of the corresponding TFLs.
- File naming: The name of each TFL output file will start with a t (for a table), f (for a figure) or l (for a listing) to denote the output type and will include an appropriate numbering as well as a reference to the associated domain and analysis set.

# 7 Analysis Sets

The following populations will be considered in the data analysis:

**Randomized population**

The Randomized population will include all subjects randomly allocated to a treatment arm.
The Randomized population will be analysed by the allocated treatment group.

**Intent-To-Treat (ITT) population**

The Intent-To-Treat (ITT) population will include all subjects randomly allocated to a treatment arm who have received at least one dose of GKT831 or placebo.
The ITT population will be analyzed by the allocated treatment group and will be the primary population for all analyses of efficacy data.

**Per Protocol (PP) population**

The Per Protocol (PP) population will include the subjects of the ITT population without a major protocol deviation.
The PP population will be regarded as a secondary population for supportive analyses of efficacy data.
The outputs planned for the PP population will be produced only if there are at least 10% fewer subjects in the PP population compared to the ITT population. That being the case, the PP population will be analyzed by the dispensed study treatment.

Protocol deviations will be defined with the classification of subjects excluded from analysis populations at the end of the study prior to unblinding.
Major protocol deviations will be summarized by the allocated treatment group.
All protocol deviations deemed important will be listed even if not leading to subjects' exclusion from the PP population.

**Safety population**

The Safety population will include all subjects who have received at least one dose of GKT831 or placebo, irrespective of whether they were randomly allocated to a treatment arm, and had at least one safety assessment.
The Safety population will be analyzed by the dispensed study treatment and will be the primary population for all analyses of safety data.

**Pharmacokinetic (PK) population**

The Pharmacokinetic (PK) population will include all subjects who have received at least one dose of GKT831, have had at least one valid PK measurement and have had no major protocol deviations relating to PK data.
The PK population will be analyzed by the received treatment group and will be the primary population for the analyses of PK data.

# 8 Methods of Analyses and Presentations

## *8.1 SUBJECT DISPOSITION*

The subject disposition summaries will be presented by allocated treatment group and overall on the randomized population.
The number and percentages of subjects in each analysis population will be summarized.
The summary of subject disposition will be showing the number and percentages of subjects belonging to the following categories:

- Subjects treated (i.e. with at least one capsule of received treatment self-administration);
- Subjects for whom the received treatment is different from the allocated treatment;
- Subjects who completed the treatment period (i.e. attended visit Week 24);
- Subjects who discontinued from the treatment period and the reasons for discontinuation;
- Subjects who completed the follow-up period (i.e. attended the follow-up visit Week 28);
- Subjects who discontinued from the follow-up period and the reasons for discontinuation.

Information on analysis populations, study completion and discontinuation will also be displayed in subject listings.

### *8.2 PROTOCOL DEVIATIONS AND/OR VIOLATIONS*

The full list of types of protocol deviations (PDvs) and their relation to the analysis sets, along with the method of identification of each PDv, are detailed in the protocol deviation criteria form which is separate to this SAP. Major PDvs are identified as 'key' deviations on the protocol deviation criteria form. This will be used as a basis for identifying subjects with PDvs throughout the study.
Protocol deviations noted during the trial by the Clinical Research Associates (CRAs), Medical Monitors or Data managers will be tracked throughout the study. In addition, PDvs will be identified programmatically in SAS® using data from the clinical database.
At the PDvs review meetings, a consolidated list of PDvs will be reviewed by the team. Prior to database lock, there will be a final review of PDvs and an agreement on the subjects final membership to or exclusion from each analysis set.
Key PDvs as well as other PDvs-related events will be summarized separately by PDv category and preferred term on the ITT population.
A listing presenting all key PDvs as well as other PDvs-related events by patient will also be provided.

### *8.3 DEMOGRAPHIC AND BASELINE CHARACTERISTICS*

The following demographic and baseline characteristics will be summarised by allocated treatment group and overall, on the ITT population:

- Age
- Age (< 65 years, ≥ 65 years)
- Sex
- Race
- Ethnicity
- Height
- Weight
- Body mass index (BMI), derived as: weight (kg)/ height(m)^2
- Child bearing potential status (females only)
- Country
- Stratification factor (disease severity level based on baseline serum GGT: < 2.5 x ULN, ≥ 2.5 x ULN)

Separately, the following disease characteristics will be summarised by allocated treatment group and overall on the ITT population:

- Duration of PBC: Time between diagnosis of BPC and randomization, derived as the number of months between the BPC diagnosis date and the randomization date:
  (date of randomization – date of BPC diagnosis)/30.4375
- PBC diagnosis age
- Liver biopsy at the BPC diagnosis date (Yes, No)

- Stage of liver fibrosis at the BPC diagnosis date (from 0 to 4)
- History of liver decompensation (Yes, No)
- Bleeding from oesophageal varices (Yes, No, Unknown)
- Upper Gastrointestinal bleed (Yes, No, Unknown)
- Ascites (Yes, No, Unknown)
- Spontaneous bacterial peritonitis (Yes, No, Unknown)
- Encephalopathy (Yes, No, Unknown)
- Baseline UDCA dose
- Duration of UDCA treatment: Time between start of UDCA treatment and randomization, derived as the number of months between the start date of UDCA treatment and the randomization date: (date of randomization – start date of UDCA treatment)/30.4375
- Baseline serum GGT
- Baseline serum ALP
- Baseline serum ALT
- Baseline serum AST
- Baseline total serum bilirubin
- Baseline serum hsCRP

In addition, listings of the above variables will be produced.

### *8.4 MEDICAL HISTORY*

Medical history is defined as a medical (or surgical) history reported in e-CRF not ongoing at the date of first dose.
Concomitant disease is defined as a medical (or surgical) history reported in e-CRF ongoing at date of first dose.
Medical histories and concomitant diseases will be coded using MedDRA.
Summaries of subjects' medical histories and concomitant diseases will be displayed by system organ class and MedDRA preferred term for each allocated treatment group and overall on the ITT population.
Listing of medical history will be produced. Also, a separate listing presenting surgical and medical procedures will be provided.

### *8.5 PRIOR AND CONCOMITANT MEDICATIONS*

Medications other than GKT137831 will be coded using the WHO-DRUG dictionary.
Medications will be defined as follows:

- Prior Medication: Any medication whose end date is before the date of first dose.
- Concomitant Medication: Any medication that started before the date of first dose and stopped on (or is ongoing after) the date of first dose OR any medication whose start date is either the same as (or after) the date of first dose.

In case of partial start and end dates, the imputation rules detailed in section 6.2 will be used.
Any medication with a missing end date will be assumed to be a concomitant medication.
Summaries of subjects' prior and concomitant medications will be displayed by INN class and preferred term for each dose group and overall on the Safety population.
In addition, a listing of prior and concomitant medications will be presented with associated flags to identify these two categories of medications.

### *8.6 STUDY DRUG EXPOSURE AND/OR COMPLIANCE*

The study drug exposure is defined as the number of days between the first dose and the last dose taken + 1 day, and it will be summarized by allocated treatment for subjects in the Safety population.
Each day during the treatment period, subjects will self-administer 4 capsules in the morning and 4 capsules in the evening.
At each dispensing visit, each subject will be given individual packs containing a number of allocated treatment bottles (70 capsules of 100 mg GKT137831 and/or matching placebo per bottle) as detailed below:

- 2 bottles at baseline/Day 1 (Visit 2) and the self-administration will be performed as follows:
  - When allocated to GKT137831 400 mg OD: 4 capsules of GKT137831 100 mg from one bottle in the morning and 4 capsules of placebo from the other bottle in the evening.
  - When allocated to GKT137831 400 mg BID: 4 capsules of GKT137831 100 mg from one bottle in the morning and 4 capsules of GKT137831 100 mg from the other bottle in the evening.
  - When allocated to placebo: 4 capsules of placebo from one bottle in the morning and 4 capsules of placebo from the other bottle in the evening
- 4 bottles at Week 2 (Visit 3): the self-administration will be performed as described above.
- 6 bottles at Week 6 (Visit 4), Week 12 (Visit 5) and Week 18 (Visit 6): the self-administration will be performed as described above.

In order to assess the compliance, the investigator or designee will count the remaining unused (i.e. returned) capsules of the received treatment at each visit.
The compliance will be recorded in the eCRF.
Overall study drug compliance is assessed as:

$$\left(\frac{\text{Total number of capsules dispensed - Total number of capsules returned}}{\text{Study drug exposure x 8}}\right)\times 100$$

The overall study drug compliance calculation assumes a subject starts dosing in the morning and finishes on their last day in the evening. Consideration of the timing (morning/evening) of the first and last dosing will be taken into account for the compliance calculation.
In addition, a compliance per visit will be listed using the same calculation.
Overall study drug compliance will be summarized using descriptive statistics by received

treatment group on the Safety population. The number (%) of subjects with compliance < 80 %, between 80% and 120% and > 120% will also be reported.
All data related to the assessment of the compliance will be listed.

## *8.7 EFFICACY DATA ENDPOINTS AND ANALYSES*

### 8.7.1 Primary Efficacy Endpoint and Analyses

The clinical response to GKT137831 is defined as the percent change from baseline to Week 24 visit in serum GGT (primary efficacy endpoint).
Absolute serum GGT values as well as absolute change and percent change from baseline to Week 24 visit in serum GGT values will be provided at baseline and Week 24 visit for each allocated treatment group using summary statistics.
Due to the small sample size, the primary analysis will be conducted using a stepwise approach.
The percent change from baseline to Week 24 visit in serum GGT will be analyzed using an Analysis of Covariance (ANCOVA) with the allocated treatment and disease severity at baseline as fixed effects, and the serum GGT baseline value as a continuous covariate.
The least square means estimate of the difference between each dose of GKT137831 and placebo alongside the 95% confidence interval will be calculated. Given two doses of GKT137831 are tested versus placebo, the 97.5% confidence interval of the least square mean of the difference between each dose of GKT137831 and placebo will also be calculated to account for the multiple testing using Hochberg adjustment.
If the normality assumption for the primary analysis is not met then the percent change from baseline in serum GGT at Week 24 will be analyzed non-parametrically through a stratified Wilcoxon Mann-Whitney (van Elteren) test. The normality assumption will be assessed through the examination of diagnostic residual plots.
The primary analysis will be performed using the ITT population and a sensitivity analysis will be done using the PP population. If the Wilcoxon Mann-Whitney test is used, a subgroup analysis for disease severity will also be performed.

### 8.7.2 Secondary and Tertiary Efficacy Endpoints and Analyses

Any analyses of the secondary and tertiary efficacy endpoints should be interpreted with care. The study has not been powered for the interpretation of these endpoints. Inferential statistical analyses performed on the secondary endpoints are included to aid interpretation and should not be considered as an alternative to the primary analysis for determining efficacy.
All the analyses on the secondary and tertiary efficacy endpoints will be performed on the ITT population.
The need for sensitivity analyses of the secondary endpoints on the PP population will be assessed after the blinded review of the major protocol deviations.

**<u>Secondary efficacy endpoints</u>**

The summaries and analyses planned for the secondary efficacy endpoints are described below.

- **Absolute and percent change in serum GGT from baseline to each post-baseline assessment.**

Absolute serum GGT values as well as absolute change and percent change from baseline in serum GGT will be provided at baseline and each post-baseline visit for each allocated treatment group using summary statistics.
The percent change from baseline to each post-baseline visit in serum GGT will be analyzed using a repeated measures ANCOVA model with the allocated treatment, the post-baseline visit and disease severity at baseline as fixed effects, the baseline serum GGT as a continuous covariate, and the interaction effect between treatment and visit.
Estimation will be performed using SAS® PROC MIXED with LSMEANS and REPEATED statement. An unstructured covariance matrix will be selected for the repeated measure analysis.
The least square means estimate of the percent change from baseline in serum GGT considered at each post baseline visit alongside its two-sided 95% CI and 97.5% CI will be presented as well as the associated standard error and the two-sided p-value.
Should the repeated measures analysis not converge or the normality assumption fail, data from each visit will be compared separately using the same method as the primary analysis.

- **Absolute change in ELF score from baseline to Weeks 12 and 24 visits**

Absolute ELF scores and absolute changes from baseline to Weeks 12 and 24 visits in ELF scores will be provided at baseline, Week 12 and Week 24 visits for each allocated treatment group using summary statistics.
The absolute change from baseline to Week 12 visit in ELF score will be analyzed using an ANCOVA with the allocated treatment and disease severity at baseline as fixed effects, and the baseline ELF score as a covariate.
The absolute change from baseline to Week 24 visit in ELF score will be analyzed similarly.

- **Absolute and percent change in serum ALP from baseline to each post-baseline assessment.**

Absolute serum ALP values as well as absolute change and percent change from baseline to each post-baseline assessment in serum ALP will be provided at baseline and each post-baseline visit for each allocated treatment group using summary statistics.
In addition, the proportion of subjects achieving a 15, 20, 30, and 40% reduction in serum ALP at each post-baseline visit will be tabulated for each allocated treatment group.
The percent change from baseline to each post-baseline visit in serum ALP will be analyzed using a repeated measures ANCOVA model with the allocated treatment, the post-baseline visit and disease severity at baseline as fixed effects, the baseline serum ALP as a continuous covariate, and the interaction effect between treatment and visit.
Estimation will be performed using SAS® PROC MIXED with LSMEANS and REPEATED statement. An unstructured covariance matrix will be selected for the repeated measure analysis.
The least square means estimate of the percent change from baseline in serum ALP considered at each post baseline visit alongside its two-sided 95% CI and 97.5% CI will be presented as well as the associated standard error and the two-sided p-value.
Should the repeated measures analysis not converge or the normality assumption fail, data from each visit will be compared separately using the same method as the primary analysis.

- **Absolute and percent change in serum levels of hsCRP and fibrinogen from baseline to each post-baseline assessment.**

Absolute serum levels of hsCRP as well as absolute change and percent change from baseline to each post-baseline assessment in serum levels of hsCRP will be provided at baseline and each post-baseline visit for each allocated treatment group using summary statistics.
The percent change from baseline to each post-baseline visit in serum level of hsCRP will be analyzed using a repeated measures ANCOVA model with the allocated treatment, the post-baseline visit and disease severity at baseline as fixed effects, the baseline serum level of hsCRP as a continuous covariate, and the interaction effect between treatment and visit.
Estimation will be performed using SAS® PROC MIXED with LSMEANS and REPEATED statement. An unstructured covariance matrix will be selected for the repeated measure analysis.
The least square means estimate of the percent change from baseline in serum level of hsCRP considered at each post baseline visit alongside its two-sided 95% CI and 97.5% CI will be presented as well as the associated standard error and the two-sided p-value.
Should the repeated measures analysis not converge or the normality assumption fail, data from each visit will be compared separately using the same method as the primary analysis.

Similar summary statistics will also be provided on absolute serum levels of fibrinogen as well as on absolute change and percent change from baseline to each post-baseline assessment.
The percent change from baseline to Week 24 visit in serum level of fibrinogen will be analyzed using an Analysis of Covariance (ANCOVA) with the allocated treatment and disease severity at baseline as fixed effects, and the baseline value of the serum level of fibrinogen as a continuous covariate.

- **Absolute and percent change in serum ALT, AST, conjugated and total bilirubin from baseline to each post-baseline assessment.**

Absolute serum ALT values as well as absolute change and percent change from baseline to each post-baseline assessment in serum ALT values will be provided at baseline and each post-baseline visit for each allocated treatment group using summary statistics.
The percent change from baseline to each post-baseline visit in serum ALT values will be analyzed using a repeated measures ANCOVA model with the allocated treatment, the post-baseline visit and disease severity at baseline as fixed effects, the baseline serum ALT value as a continuous covariate, and the interaction effect between treatment and visit.
Estimation will be performed using SAS® PROC MIXED with LSMEANS and REPEATED statement. An unstructured covariance matrix will be selected for the repeated measure analysis.
The least square means estimate of the percent change from baseline in serum ALT considered at each post baseline visit alongside its two-sided 95% CI and 97.5% CI will be presented as well as the associated standard error and the two-sided p-value.
Should the repeated measures analysis not converge or the normality assumption fail, data from each visit will be compared separately using the same method as the primary analysis.

Serum AST and serum level of total bilirubin will be summarized and analysed similarly.

Similar summary statistics will also be provided on absolute serum level of conjugated bilirubin as well as on absolute change and percent change from baseline to each post-baseline assessment.
The percent change from baseline to Week 24 visit in serum level of conjugated bilirubin will be analyzed using an Analysis of Covariance (ANCOVA) with the allocated treatment and disease severity at baseline as fixed effects, and the baseline value of the serum level of conjugated bilirubin as a continuous covariate.

- **Absolute and percent change in the FIB-4 and APRI scores from baseline to each post-baseline assessment**

Absolute FIB-4 scores as well as absolute change and percent change from baseline to each post-baseline assessment in FIB-4 scores will be provided at baseline and each post-baseline visit for each allocated treatment group using summary statistics.
The percent change from baseline to Week 24 visit in FIB-4 score will be analyzed using an Analysis of Covariance (ANCOVA) with the allocated treatment and disease severity at baseline as fixed effects, and the baseline value FIB-4 score of as a continuous covariate.

The APRI scores will be summarized and analysed similarly.

- **Absolute and percent change in liver stiffness from baseline to Week 24 visit.**

Only subjects with liver stiffness values at both baseline and Week 24 visits will be taken into account for the summaries and analyses.
Absolute liver stiffness values as well as absolute change and percent change from baseline to Week 24 visit in liver stiffness values will be provided at baseline and Week 24 visit for each allocated treatment group using summary statistics.
The percent change from baseline to Week 24 visit in liver stiffness will be analyzed using an ANCOVA with the allocated treatment and disease severity at baseline as fixed effects, and the baseline liver stiffness value as a continuous covariate.

- **Absolute and percent change in serum levels of collagen fragments (ProC3, ProC5, C3M, C4M, and BGM) from baseline to Weeks 12 and 24 visits.**

Absolute serum levels of collagen fragments and absolute changes from baseline to Week 12 and Week 24 visits in serum levels of collagen fragments will be provided at baseline, Week 12 and Week 24 visits for each allocated treatment group using summary statistics.
The percent change from baseline to Week 12 visit in serum level of collagen fragments will be analyzed using an ANCOVA with the allocated treatment and disease severity at baseline as fixed effects, and the baseline serum level of collagen fragments as a continuous covariate.
The percent change from baseline to Week 24 visit in serum level of collagen fragments will be analyzed similarly.

- **Absolute change in Quality of Life scores (based on the 6 domains of the PBC-40 Questionnaire) and Pruritus score (based on Pruritus VAS) from baseline to Weeks 12 and 24 visits.**

The PBC-40 questionnaire will be summarized by domain using the scoring described in the coded-PBC-40 document.

If data are missing from a domain (typically missed or duplicated answers), the whole domain should be discarded if <50% of the items are completed. If >50% of the items have responses then the median value for the completed items in the domain should be ascribed to the missing items.

For each of the 6 domains (symptoms, itch, fatigue, cognition, social, emotional), the total scores as well as the absolute and percent changes from baseline to Week 12 and Week 24 visits in the domain's total scores will be provided at baseline, Week 12 and Week 24 visits for each allocated treatment group using summary statistics.

The Pruritus scores will be summarized similarly.

**Tertiary efficacy endpoints**

The summaries and analyses planned for the tertiary efficacy endpoints are described below.

- **Absolute and percent change in serum level of total bile acids from baseline to Weeks 12 and 24 visits.**

Absolute serum level of total bile acids and absolute changes from baseline to Weeks 12 and 24 visits in serum level of total bile acids will be provided at baseline, Week 12 and Week 24 visits for each allocated treatment group using summary statistics.

The percent change from baseline to Week 12 visit in serum level of total bile acids will be analyzed using an ANCOVA with the allocated treatment and disease severity at baseline as fixed effects, and the baseline serum level of total bile acids as a covariate.

The percent change from baseline to Week 24 visit in serum level of total bile acids will be analyzed similarly.

- **Proportion of subjects who meet the definition of PBC responder** applying the following two criteria:
  - ALP < 1.67xULN
  - ALP < 1.67xULN, ALP reduction ≥ 15%, and total bilirubin < ULN (composite criterion)

For each of the two above PBC response criteria, the proportion of responders at Weeks 12 and 24 visits will be presented for each allocated treatment group.

### 8.7.3 Exploratory Efficacy Endpoints and Analyses

The below planned analyses on the exploratory efficacy endpoints will be performed only on the ITT subjects who meet the following two criteria:

- A sample of the exploratory efficacy parameter is available at Baseline (V2) visit
- At least one follow-up sample of the exploratory efficacy parameter is available either at Week 12 (V5) visit or at Week 24 (V7) visit

- **Absolute and percent change in serum C4 and FGF19 from baseline to Weeks 12 and 24 visits.**

Absolute serum C4 values as well as absolute and percent changes from baseline to Week 12 and Week 24 visits in serum C4 values will be provided at baseline, Week 12 and Week 24 visits for each allocated treatment group using summary statistics.
FGF19 values will be summarized similarly.

- **Absolute and percent change in serum IL-6, CK-18 M30, and CK-18 M65, from baseline to Weeks 12 and 24.**

The summaries on serum IL-6 and CK-18 M30 and M65 fractions will be similar to those above-described for serum C4.

- **Absolute and percent change in serum IgM, IL-4, IL-12, IL-17A, and interferon γ, from baseline to Weeks 12 and 24.**

The summaries on serum IgM, IL-4, IL-12, IL-17A, and interferon γ will be similar to those above-described for serum C4.

- **Assessment of additional biomarkers of interest.**

If relevant, data-driven analyses on additional biomarkers of interest will be performed. If performed, these post-hoc analyses will be described in a SAP Change Log and the results will be included in the exploratory efficacy analyses.

- **Assessment of metabolomics signatures.**

If relevant, metabolomics analyses will be performed. If performed, these analyses will be outlined in a separate analysis plan and reported in a dedicated study report included as an appendix to the clinical study report.

- **Assessment of pharmacogenetics.**

If relevant, pharmacogenetics analyses of interest will be performed. If performed, these analyses will be outlined in a separate analysis plan and reported in a dedicated study report included as an appendix to the clinical study report.

### 8.7.4 Subgroup Analysis

Subgroup summaries/analyses will be performed using the following subgroups of interest:

- Subgroups of interest 1: subjects with GGT values below and above the stratification cut off value (≥2.5xULN) (for the primary efficacy endpoint analysis only)
- Subgroups of interest 2: subjects with hsCRP value below and above the ULN at Baseline (for the analysis of hsCRP only).

- Subgroups of interest 3: subjects with ELF score value below 9.8 at Baseline and subjects with ELF score value equal to or greater then 9.8 at Baseline (for the analysis of ELF scores only).
- Subgroups of interest 4: subjects with liver stiffness value below 9.6 kPa at Baseline and subjects with liver stiffness value equal to or greater than 9.6 kPa at Baseline (for the analysis of liver stiffness only).
- Subgroups of interest 5: subjects with any Baseline pruritus VAS value and subjects with a Baseline pruritus VAS value greater than 0.

For each subgroup of interest, any subject with a missing/unknown result will be excluded from the subgroup summaries/analyses.
Subgroup summaries/analyses will only be performed for a minimum subgroup size of at least 20% of the analyzed population, unless otherwise specified.
For a given endpoint, the subgroup summaries/analyses will be run on each of the analysis populations considered for the global summaries/analyses unless otherwise specified.

### *8.8 PHARMACOKINETIC/ PHARMACODYNAMIC ENDPOINTS AND ANALYSES*

Plasma concentrations of GKT137831 and its main phase 1 metabolite GKT138184 will be summarized using descriptive statistics by received dose level and planned visit. Any plasma concentrations below the limit of quantification will be considered as 0 for the calculation of the summary statistics.
The actual sampling time will be presented in listings. Pre-dose samples collected post dose will automatically be flagged for exclusion in the summary statistics.
Samples will be analyzed to determine plasma drug concentrations and thus to aid investigation of any PK/PD relationship with PD and/or efficacy endpoints, and optionally for the relationships between plasma drug concentrations and pharmacogenomics data.
A population PK model describing the plasma concentrations of GKT137831 and GKT138184 will be developed using non-linear mixed-effects modelling.
Relationships between drug concentrations/exposure measures and selected PD and/or therapeutic efficacy endpoints in the same subject will be graphically explored and formal PK/PD (exposure-response) analyses may be performed.
A separate Modeling and Simulation Analysis Plan (MSAP) for the PK/PD modeling describing the general approach to be taken will be finalized prior to database lock. The actual execution of any PK/PD modeling will depend upon the data, and full details of this will be provided in a separate report, which will be appended to the clinical study report (CSR).

### *8.9 QUALITY OF LIFE OR PHARMACOECONOMIC ENDPOINTS AND ANALYSES*

For the Quality of life questionnaire see subsection 'Tertiary efficacy endpoints' in Section 8.7.2.

### *8.10 SAFETY DATA ENDPOINTS AND ANALYSES*

All Safety analyses will be performed on the Safety Population.

### 8.10.1 Adverse Events (AEs)

All AEs will be coded using MedDRA dictionary. The version of the dictionary will be provided in the adverse events tables, figures and listing footnotes.
All information on AEs will be listed by subject.
Treatment-Emergent Adverse Events (TEAEs) will be defined as AEs that started on or after the date of the first dose of the received treatment.
AEs that occurred prior to the first dose of the received treatment and increased in severity and/or relationship after the first dose of the received treatment will also be regarded as TEAEs.
If it is not possible to determine whether an AE is a TEAE due to a missing start date, the AE will be regarded as a TEAE unless the stop date indicates that the AE occurred prior to the first dose of the received treatment.
All reported TEAEs will be summarized.
The number and percentage of subjects experiencing the following categories of AEs will be summarized for each received treatment group and for all received treatment groups combined both overall and by system organ class and preferred term:

- TEAEs
- Related TEAEs
- Serious TEAEs
- Related and serious TEAEs
- TEAEs leading to study drug discontinuation
- TEAEs leading to study drug interruption
- Non-TEAEs

In addition, the number and percentage of subjects experiencing the following categories of AEs will be summarized for each received treatment group and for all received treatment groups combined by system organ class, preferred term and maximal severity:

- TEAEs
- Related TEAEs

For all system organ classes and preferred terms with a TEAE incidence rate of at least 5% over all received treatment groups combined, the number and percentage of subjects will be summarized for each received treatment group and for all received treatment groups combined by system organ class and preferred term.
Subjects will be counted only once within each system organ class, preferred term and severity.

- If a subject has experienced more than one adverse events coded within the same system organ class then this subject will be counted only once within that system organ class.
- If a subject has experienced more than one adverse events coded to the same preferred term within a system organ class then the subject will be counted only once for that preferred term and within that system organ class.
- If a subject has experienced more than one adverse events coded to the same preferred term within a system organ class then the subject will be counted only once under the

maximum severity for that preferred term and within that system organ class. The maximal severity is defined as first death (the highest possible level of severity), then life-threatening followed by severe, moderate, and mild (the lowest possible level of severity).

If a subject has experienced any AE with a missing severity then the AE's grade will be imputed as the maximal reported grade over all AEs experienced by that subject (excluding death).
If a subject has experienced any AE with a missing relationship then the AE's relationship will be assumed to be related to study drug.

**Adverse Events of Special Interest (AESI)**

Potential effect of GKT137831 on arterial blood pressure and its clinical relevance will be evaluated. Systolic and diastolic blood pressure values will be examined together with temporally associated blood pressure related events including adverse events and serious adverse events, changes in study drug administration and other related changes in concomitant medications.
Drug induced liver injury related events will be considered as AESI as well.
In addition, the medical history and other possible contributing risk factors will be assessed.
The following categories of AEs will be listed by subject:

- All adverse events
- All TEAEs
- All serious AEs
- All AEs leading to study drug discontinuation
- All AEs leading to study drug interruption
- All AEs leading to death
- All AESI

### 8.10.2 Clinical Laboratory Evaluations

Laboratory parameters (hematology, biochemistry, TSH levels, urinalysis) results and changes from baseline will be summarized with descriptive statistics by visit for each received treatment group and for all received treatment groups combined.
The changes from baseline will also be displayed by visit for each received treatment group using box plots.
For each laboratory parameter, the incidence of abnormal results (i.e. outside of the normal range) will be summarized by visit for each received treatment group and for all received treatment groups combined. For dipstick parameters with a categorical response, any positive result will be considered abnormal.
For each parameter, the shift from baseline to each planned post-baseline visit in the range of values (low, normal, high and abnormal [low/high]) will be summarized by post-baseline visit with the number and percentages of subjects.
For the calculation of descriptive statistics, all laboratory results will be converted to SI units. Differential counts will be presented both as absolute values and as percentages of white blood cells.
The NCI-CTCAE criteria v4.0 [3] will be used to determine severity.

For each hematological parameter, the number and percentage of the worst NCI-CTCAE grades will be presented for each received treatment group. The shift from baseline grade to the worst NCI-CTCAE grade will also be presented by received treatment group for each hematological parameter.
For each biochemistry parameter, the same summaries on the worst NCI-CTCAE grades will be produced.
All data will be listed in subject data listings and abnormal results will be flagged.

The laboratory parameters assessments reported as '<LOQ' will be imputed with and displayed as the calculated value equal to LOQ/2 in the listings.

### 8.10.3 Vital Signs

Vital signs include Pulse rate, Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP).
Pulse rate, SBP and DBP values as well as change from baseline to each planned visit will be summarized with descriptive statistics by visit for each received treatment group and for all received treatment groups combined.
The shift from baseline to each planned post-baseline visit for each of the criteria specified in Table 2 will be summarized by post-baseline visit for each received treatment group and for all received treatment groups combined.

**Table 2: Vital sign values of clinical importance**

| Vital Sign | Criteria | Flag |
|---|---|---|
| Pulse Rate | < 55 bpm | Low (L) |
| | ≥ 55 to ≤ 100 bpm | Normal |
| | > 100 bpm | High (H) |
| Systolic Blood Pressure | < 90 mmHg | Low (L) |
| | ≥ 90 to ≤ 140 mmHg | Normal |
| | > 140 mmHg | High (H) |
| Diastolic Blood Pressure | < 45 mmHg | Low (L) |
| | ≥ 45 to ≤ 90 mmHg | Normal |
| | > 90 mmHg | High (H) |

Box plots will be produced to plot the vital signs change from baseline at each visit.
Vital signs (including Body Weight) will be listed by subject and visit with both conventional and Standard International values and all results considered abnormal as per criteria defined in Table 2 will be flagged.

### 8.10.4 ECG

ECG parameters include Heart Rate, QRS axis, PR interval, RR interval, QRS duration and QT interval.

In addition, QTcB interval and QTcF interval will be calculated based on the following formulae:

QTcB = QT interval/Square root of RR interval
QTcF = QT interval/Cubed root of RR interval

For all the above ECG parameters, the values as well as the changes from baseline to each planned visit will be summarized with descriptive statistics by visit for each received treatment group and for all received treatment groups combined.
For all the above ECG parameters, the number and percentage of subjects with abnormal results will be presented by visit for each received treatment group and for all received treatment groups combined.
In addition, the shift from baseline to each planned post-baseline visit in the overall ECG interpretation (normal, abnormal not clinically significant and abnormal clinically significant) will be summarized by post-baseline visit for each received treatment group and for all received treatment groups combined.
ECG parameters values will be listed by subject and visit and all results considered abnormal as per criteria defined in Table 3 will be flagged.

**Table 3: Abnormal ECG -Criteria**

| QTc Interval | Criteria (msec) | Flag |
|---|---|---|
| Change from Baseline | ≤ 30 | Low (L) |
| | > 30 to ≤ 60 | Mid (M) |
| | > 60 | High (H) |
| Actual Value | > 450 to ≤ 480 | Low (L) |
| | > 480 to ≤ 500 | Mid (M) |
| | > 500 | High (H) |

For QTcB interval and QTcF interval, the categorized values at each visit and the categorized changes from baseline to each post-baseline visit will be summarized by post-baseline visit with the number and percentage of subjects for each received treatment group and for all received treatment groups combined. For both parameters, the categorized worst change from baseline over all planned post-baseline visits will be summarized similarly.
Box plots will be produced for each ECG parameter (Heart Rate, QRS axis, PR interval, RR interval, QRS duration, QT interval, QTcB interval and QTcF interval) showing change from baseline at each visit.
All ECG parameters values will be listed by subject and visit and all results considered abnormal as per criteria defined in Table 3 will be flagged.

# 9 Interim Analyses

An interim analysis is planned once at least 80-90% of the planned number of subjects to be randomized in the study have completed their Week 6 visit.
Only data collected for these subjects, prior to and including the date of the Week 6 visit for the last randomized subject included in the interim analysis cohort, will be taken into consideration for the interim analysis.

The purpose of the interim analysis is to support decision-making regarding the further development of GKT137831. It is not intended to amend the study protocol or stop the trial due to futility or overwhelming efficacy.
The primary efficacy endpoint of the interim analysis is the percent change from baseline to Week 6 visit in serum GGT.
The interim analysis will only be run on this primary efficacy endpoint and on the following secondary efficacy endpoints (no other endpoint will be analysed):

- Absolute change from baseline to Week 6 visit in serum GGT
- Absolute and percent change from baseline to Week 6 visit in serum levels of ALP
- Absolute and percent change from baseline to Week 6 visit in serum levels of hsCRP and fibrinogen
- Absolute and percent change from baseline to Week 6 visit in serum levels of ALT, AST, conjugated and total bilirubin.
- Absolute and percent change from baseline to Week 6 visit in the FIB-4 and APRI scores.

The interim analyses will also be run for the following 4 subgroups of interest:

- Subgroups of interest 1: subjects with GGT values below and above the stratification cut off value (≥2.5 x ULN) (for the primary efficacy endpoint of the interim analysis only).
  If only very few subjects have a Baseline GGT below 2.5 x ULN then the stratification cut off value for the analysis will be the median GGT at Baseline.
- Subgroups of interest 2: subjects below and above 65 years of age (for the primary efficacy endpoint of the interim analysis only)
- Subgroups of interest 3: subjects with ALP value below and above median at Baseline (for the primary efficacy endpoint of the interim analysis only)
- Subgroups of interest 4: subjects with hsCRP value below and above the ULN at Baseline (for the analysis of hsCRP only)

Subgroup summaries/analyses will only be performed for a minimum subgroup size of at least 20% of the interim analysis population, unless otherwise specified.

In addition, the gender and the age as well as the serum GGT, the serum levels of ALP, ALT, and hsCRP values will be summarized at Baseline for each allocated treatment group.

The interim analysis will be conducted by the unblinded SMB statistician, who has no involvement with the study conduct. The interim analysis outputs will be communicated to the Sponsor's Chief Medical Officer.
The interim analysis described above does not include assessments of safety data. The safety data will be reviewed periodically and separately as described in the study SMB charter.

# 10 Changes to Planned Analyses

The final analysis will be performed according to a 2-stage process detailed in Section 6.1.

# 11 Document History

| Date | Version | Modified by | Brief details of changes made to template |
|---|---|---|---|
| 15OCT2018 | 1.0 | Jean-Luc Beffy | Initial final version |
| 15MAR2019 | 2.0 | Jean-Luc Beffy | Updated final version<br>The main changes are the following:<br>1. Addition of Section 6.1 to clarify the timing of the final analysis.<br>2. Addition of Appendix 1 to provide the list of Top Line tables. |

# 12 References

[1] New Drug Application (NDA) 207999 submitted by Intercept Pharmaceuticals, Inc. for obeticholic acid (OCA) proposed for the treatment of Primary Biliary Cirrhosis (PBC), Gastrointestinal Drug Advisory Committee (GIDAC) Meeting, April 7th, 2016.

[2] Hirschfield GM et al. Efficacy of Obeticholic Acid in Patients With Primary Biliary Cirrhosis and Inadequate Response to Ursodeoxycholic Acid. Gastroenterology 2015; 148: 751-761.

[3] CTCAE v4.03:

http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

# 13 Appendix 1: List of Top Line tables

| Table order | Table number | Table title |
|---|---|---|
| 1 | 14.1-1 | Subject Disposition (Randomized Population) |
| 2 | 14.1-3 | Summary of Demographic and Baseline Characteristics (ITT Population) |
| 3 | 14.1-5.1 | Disease Characteristics (ITT Population) |
| 4 | 14.2-1.1.1 | Summary of Absolute and Percent Change from Baseline to Week 24 Visit in Serum GGT (ITT Population) |
| 5 | 14.2-1.1.2 | Analysis of Percent Change from Baseline to Week 24 Visit in Serum GGT (ITT Population) |
| 6 | 14.2-1.3.1.1 | Summary of Percent Change from Baseline to Week 24 Visit in Serum GGT by Subgroup of Baseline GGT values (ITT Population) |
| 7 | 14.2-1.3.1.2 | Analysis of Percent Change from Baseline to Week 24 Visit in Serum GGT by Subgroup of Baseline GGT values (ITT Population) |
| 8 | 14.2-2.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in Serum GGT (ITT Population) |
| 9 | 14.2-2.1.2 | Analysis of Percent Changes from Baseline in Serum GGT across all Post Baseline Visits [a] Model Estimates (ITT Population) |
| 10 | 14.2-4.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in Serum Level of ALP (ITT Population) |
| 11 | 14.2-4.1.2 | Analysis of Percent Change from Baseline to each Post-Baseline Visit in Serum Level of ALP Model Estimates (ITT Population) |
| 12 | 14.2-5.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in Serum Level of hsCRP (ITT Population) |
| 13 | 14.2-5.1.2 | Analysis of Percent Change from Baseline to each Post-Baseline Visit in Serum Level of hsCRP Model Estimates (ITT Population) |
| 14 | 14.2-7.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in Serum Level of ALT (ITT Population) |
| 15 | 14.2-7.1.2 | Analysis of Percent Change from Baseline to each Post-Baseline Visit in Serum Level of ALT Model Estimates (ITT Population) |
| 16 | 14.2-8.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in Serum Level of AST (ITT Population) |

| Table order | Table number | Table title |
|---|---|---|
| 17 | 14.2-9.2.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in Serum Level of Total Bilirubin (ITT Population) |
| 18 | 14.2-10.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in FIB-4 score (ITT Population) |
| 19 | 14.2-10.1.2 | Analysis of Percent Change from Baseline to Week 24 Visit in FIB-4 score (ITT Population) |
| 20 | 14.2-11.1.1 | Summary of Absolute and Percent Change from Baseline to each Post-Baseline Visit in APRI score (ITT Population) |
| 21 | 14.2-11.1.2 | Analysis of Percent Change from Baseline to Week 24 Visit in APRI score (ITT Population) |
| 22 | 14.2-12.1.1 | Summary of Change from Baseline to Week 24 Visit in Liver Stiffness (ITT Population) |
| 23 | 14.2-12.1.2 | Analysis of Change from Baseline to Week 24 Visit in Liver Stiffness (ITT Population) |
| 24 | 14.2-12.2.1 | Summary of Change from Baseline to Week 24 Visit in Liver Stiffness by Baseline Liver Stiffness Value Subgroup (ITT Population) |
| 25 | 14.2-12.2.2 | Analysis of Change from Baseline to Week 24 Visit in Liver Stiffness by Baseline Liver Stiffness Value Subgroup (ITT Population) |
| 26 | 14.2-13.2.1.1 | Summary of Change from Baseline to Week 24 Visit in Serum Level of Collagen Fragments (ITT Population) |
| 27 | 14.2-13.2.1.2 | Analysis of Change from Baseline to Week 24 Visit in Serum Level of Collagen Fragments (ITT Population) |
| 28 | 14.2-15.2 | Summary of Change from Baseline to Week 24 Visit in each Domain Score of the PBC-40 Questionnaire (ITT Population) |
| 29 | 14.2-16 | Summary of Change from Baseline to Week 12 and Week 24 Visits in Pruritus Score (VAS) (ITT Population) |
| 30 | 14.2-17 | PBC Responders Rates at Week 12 and 24 Visits Applying Two Response Criteria: Serum ALP < 1.67xULN and Combined Criterion Serum ALP < 1.67xULN, ALP Reduction ≥ 15%, and total bilirubin < ULN (ITT Population) |