1.0 Title page

Statistical Analysis Plan

Study M18-803

A Prospective, Open-Label, Single-Arm, Phase 2, Multicenter Study Evaluating the Efficacy of Venetoclax Plus Ibrutinib in Subjects with T-Cell Prolymphocytic Leukemia

Date: 18 Jul 2019

Version 2.0

2.0 Table of Contents

List of Tables

3.0 Introduction

This statistical analysis plan (SAP) describes the full (safety and efficacy) statistical analyses for venetoclax (ABT-199) Protocol M18-803 dated May 02, 2019. It will provide details of statistical methods to be followed for the study. More details in terms of analysis conventions to guide the statistical programming work will be provided in SPP (Statistical Programming Plan).

Unless noted otherwise, all analyses will be performed using SAS® version 9.3 or later (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

SAP will not be updated in case of future administrative or minor amendments to the protocol unless the changes have any impact on the analysis of the study data.

4.0 Study Background

4.1 Objective

Primary Objective:

The primary objective of this study is to evaluate the efficacy of venetoclax in combination with ibrutinib in subjects with relapsed or refractory (R/R) T-cell prolymphocytic leukemia (T-PLL). Efficacy will be measured by overall response rate (ORR) in R/R subjects.

Secondary Objectives:

- Secondary efficacy analyses will be performed by the following endpoints:
  - Progression-free survival (PFS)
  - Duration of response (DoR)
  - Time-to-progression (TTP)
  - Event-free survival (EFS)
  - Overall survival (OS)

- To assess safety and tolerability in subjects with T-PLL treated with venetoclax in combination with ibrutinib. Safety will be assessed by
  - Adverse event (AE) monitoring, physical examinations, vital sign measurements, electrocardiogram variables and clinical laboratory testing.
- To assess the ability of the subset of transplant-naïve subjects to proceed to further autologous/allogeneic stem cell transplantation, will be measured by
  - Total number of eligible subjects reaching autologous or allogeneic transplantation.

Exploratory Objective:

The exploratory objective is to identify biomarker/genetically-defined subgroups regarding response and survival. Types of biomarkers may include nucleic acids, proteins, lipids, or metabolites. These biomarker assessments may be used to assess and generate prognostic, predictive, pharmacodynamics, or surrogate biomarker signatures. Analyses of biomarkers may be done as an external study report and will not be a part of the clinical study report.

4.2 Study Design

This is a Phase 2, open-label, single-arm, multicenter study designed to evaluate the efficacy and safety of venetoclax when administered with ibrutinib, in subjects with R/R T-PLL.

4.2.1 Study Design and Design Diagram

Approximately 37 subjects with T-PLL will be enrolled based on the optimal adaptive 2-stage design.[1]

- Stage 1: Enroll 14 subjects with R/R T-PLL and move to Stage 2 if 4 or more responders as follows:
  - response assessment for Stage 1 will be performed on a continued basis until all 14 subjects have enrolled into Stage 1 and have completed the Week 24 disease assessment.

- Stage 2: Enroll up to an additional 23 subjects with T-PLL in Stage 2 (up to a total of 37 subjects for Stage 1 and Stage 2) depending on the number of R/R T-PLL responders at Stage 1.

Interim analysis will occur after all 14 R/R T-PLL subjects enrolled in Stage 1 of the study have completed their Week 24 disease assessments. The study will stop for futility if the number of responders at Stage 1 is ≤ 3. If 4 or more subjects respond prior to enrolling all 14 subjects in Stage 1, enrollment will not pause between stages. The maximal sample size will be 37 subjects if number of responders at Stage 1 is 4. If there are 7 or more responders out of 14 subjects in Stage 1, 17 R/R subjects will be enrolled in Stage 2 (see Section 4.4 for details).

Subjects who are treatment-naïve (i.e., unsuitable for alemtuzumab or alemtuzumab treatment is unavailable) may be enrolled during Stage 2 of the study. The number of treatment-naïve subjects to be enrolled will depend on the required sample size for the R/R subjects in Stage 2 which is based on the number of responders from Stage 1 and considering the maximum of 37 subjects in total to be enrolled in the study.

It is to be noted that the opening of Stage 2 will be determined based on R/R T-PLL subjects, enrolled in Stage 1.

Safety will be continuously assessed and dose adjustment or tumor lysis syndrome (TLS) risk management will be considered for both venetoclax and ibrutinib.

4.2.2 Variables Used for Stratification at Randomization

This is an open-label, single-arm study. There is no randomization for this study. All subjects will be assigned a unique identification number by the IRT at the screening visit.

4.3 Endpoint

4.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint in this study is ORR in R/R T-PLL subjects.

4.3.2 Secondary Efficacy Endpoints

The following are the secondary efficacy endpoints in the study: PFS, DoR, TTP, EFS, OS, total number of eligible subjects (transplant-naïve subjects who have achieved CR) reaching autologous or allogeneic transplantation.

4.3.3 Safety Endpoint

Safety data of the study will be assessed through adverse events reporting, including adverse event (AE) monitoring, laboratory values, physical examinations, vital signs, and ECG assessment. Other protocol-specified tests that are deemed critical to the safety will also be analyzed. (See Section 7.0 for analysis details).

4.3.4 Pharmacological Endpoint

Plasma samples for venetoclax and ibrutinib will be collected to determine trough concentrations (C_{trough}). A nonlinear mixed effects modeling approach may be used to estimate the population central values and the individual values of venetoclax parameters such as clearance.

4.4 Sample Size Justification

Approximately 37 subjects with T-PLL will be enrolled based on the optimal adaptive 2-stage design. In Stage 1, 14 subjects with R/R T-PLL will be enrolled and the number of responders in Stage 1 will be used to determine the number of subjects to be enrolled for Stage 2.

Sample size calculation is done based on clinical hypothesis (using optimal adaptive 2-Stage design considering 1-sided significance level of 0.05 and power of 80%): In subjects with R/R T-PLL, the treatment with venetoclax plus ibrutinib will be considered promising if the combination yields a best ORR of 40% or greater, which is the alternative hypothesis. A poor rate is defined as best ORR of 20% or less, which is the null hypothesis.

The details of the design are presented in the Table 1 below:

Table 1. Sample Size Using Optimal Adaptive 2-Stage Design

H_0: $p \leq 20\%$ versus H_1: $p \geq 40\%$ 1-β = 80%, 1-sided test, significance level = 0.05				Comments
$n_1 = 14$				
S	**$n_2(S)$**	**n(S)**	**r(S)**	
≤ 3	0	14	0	Stop for futility after Stage 1
4	23	37	11	R/R versus treatment-naïve Stage 2 sample size distribution will vary depending on the number of responses from Stage 1
5	20	34	10	
6	20	34	10	
≥ 7[a]	17	31	9	

n_1: Sample size for Stage 1.

S: Number of responses from Stage 1.

$n_2(S)$: Sample size for Stage 2.

n(S): Total sample size of Stage 1 and Stage 2.

r(S): Critical value of total number of responses from Stage 1 and Stage 2 to demonstrate efficacy.

a. The Stage 2 portion will still be conducted to evaluate efficacy in additional subjects even if the Stage 1 portion shows significant efficacy.

4.5 Interim Analysis

A planned interim analysis for efficacy will be conducted after 14 subjects with R/R T-PLL enrolled in Stage 1 have completed their Week 24 disease assessments. The study will stop for futility if the number of responders at Stage 1 is ≤ 3. If 4 or more subjects respond prior to enrolling all 14 subjects in Stage 1, enrollment will not pause between stages. The maximal sample size will be 37 subjects if number of responders at Stage 1 is 4. If there are 7 or more responders out of 14 subjects in Stage 1, 17 R/R subjects will be enrolled in Stage 2.

4.6 Multiplicity Testing Procedures for Type-I Error Control

This is a Phase II hypothesis generating study; therefore, no multiplicity adjustment will be applied for the statistical analysis.

4.7 Missing Data Imputation

The missing data imputation will not be performed for this study. For time-to-event endpoints, censoring will be defined for missing data.

5.0 Analysis Populations and Important Subgroups

5.1 Analysis Population

The Full Analysis Set (FAS) includes all R/R subjects who received at least 1 dose of study drug (either venetoclax or ibrutinib). The FAS will be used for all safety, pharmacokinetic, efficacy and baseline analyses.

The analyses for the treatment naïve subjects may be conducted for similar efficacy and safety endpoints, if needed, otherwise data will be listed.

5.2 Subgroup

Subgroups will be based on demographic and baseline characteristics (age, gender, race, and high risk T-PLL factors such as tumor burden at screening; and other genetical abnormalities.). Full details will be provided in the SPP.

6.0 Efficacy Analyses

6.1 General Considerations

Efficacy will be evaluated based on clinical response criteria adapted from the response assessment criteria proposed by the international T-PLL study group. Confirmation of responses CR and CRi will be performed based on modified RECIL 2017, which is based on CT imaging (or MRI).

After the data collection is completed and reviewed for completeness and all data management quality assurance (QA) and quality control (QC) procedures are performed, the clinical database data will be extracted for documentation and statistical analyses of

the efficacy and safety data for the pre-specified cutoff date. Other considerations will be detailed in the SPP, if needed.

6.2 Primary Efficacy Analysis

Overall Response Rate (ORR) (CR + CRi + PR) will be defined as the proportion of subjects who achieved best response of confirmed complete remission (CR), confirmed complete remission with incomplete marrow recovery (CRi), or partial remission (PR) based on the clinical response criteria proposed by the international T-PLL study group as assessed by investigator up through the Week 24 disease assessment (including early study termination). The corresponding 95% confidence interval for the proportion using Binomial distribution will be constructed. Subjects who have not achieved any component of ORR will be considered as nonresponders in the calculation of ORR at the time of analysis. An interim efficacy analysis will be performed when 14 subjects in Stage 1 have completed the Week 24 disease assessments. The final primary efficacy analysis will be performed for Stage 1 and Stage 2 subjects after their completion of at least Week 24 disease assessments.

6.3 Secondary Efficacy Analysis

The secondary efficacy analysis will be performed for the following endpoints:

Progression-Free Survival (PFS) will be defined as the number of days from the date of first dose of study drug to the date of earliest disease progression or death. All events of disease progression will be included regardless whether the event occurred while the subject was taking the study drug or had previously discontinued the study drug. If the subject has not experienced disease progression or death, the subject's data will be censored at the date of last disease assessment. Data for subjects who receive non-protocol specified T-PLL therapy (i.e., excluding stem cell transplantation for eligible CR subjects) prior to disease progression will be censored at the last disease assessment prior to receiving non-protocol therapy. Data for subjects without any disease assessments performed after the baseline visit will be censored at the time of dosing plus 1 day. PFS

will be analyzed by Kaplan-Meier methodology. Median PFS will be calculated and 95% confidence interval for median PFS will be presented.

Duration of Response (DoR) will be defined as the number of days from the date of first response (CR, CRi, or PR) to the earliest date that progressive disease (PD) or death. For subjects who have a PR before CR, or CRi, in subsequent visits, the DoR is computed from the earliest PR. If the subject has not experienced disease progression or death, the subject's data will be censored at the date of last disease assessment. For subjects who do not have subsequent disease assessment after their first response, the subject data will be censored at the first response. For subjects who never experience response, the subject's data will not be included in the analysis. DoR will be analyzed by Kaplan-Meier methodology. Median DoR will be calculated and the corresponding 95% confidence interval will be presented.

Time- to- Progression (TTP) will be defined as the number of days from the date of first dose of study drug to the date of earliest disease progression. All events of disease progression will be included regardless whether the event occurred while the subject was taking the study drug or had previously discontinued the study drug. If the subject has not experienced disease progression, the subject's data will be censored at the date of last disease assessment. Data for subjects who receive non-protocol specified T-PLL therapy (i.e., excluding stem cell transplantation for eligible CR subjects) prior to disease progression will be censored at the last disease assessment prior to receiving non-protocol therapy. Data for subjects without any disease assessments performed after the baseline visit will be censored at the time of dosing plus 1 day. TTP will be analyzed by Kaplan-Meier methodology. Median TTP will be calculated and 95% confidence interval for median TTP will be presented.

Event-free survival (EFS) will be defined as the number of days from the date of first dose of study drug to the date of earliest disease progression, death, or start of a new anti-T-PLL therapy. If the specified event (disease progression, death, start of a new anti-T-PLL treatment) does not occur, patients will be censored at the date of last disease assessment or date of stem cell transplantation. Data for subjects without any disease assessments

performed after the baseline visit will be censored at the date of first dose plus 1 day. EFS will be analyzed by Kaplan-Meier methodology. EFS will be calculated and 95% confidence interval for median EFS will be presented.

Overall Survival (OS) will be defined as the number of days from the date of first dose of study drug to the date of death due to any cause. All events of death will be included, regardless of whether the event occurred while the subject was still taking any study drug, or after the subject discontinued study drug. For subjects who did not die, their data will be censored at the date of last study visit or the last known date to be alive, whichever is later. OS will be analyzed by Kaplan-Meier methodology. Median OS will be calculated and 95% confidence interval for median OS will be presented. Data from subjects that are alive at the time of analysis will be censored at the date of last study visit or the last known date the subject was alive, whichever is later.

6.4 Sensitivity Analysis:

For all the time-to-event endpoints (PFS, TTP, DoR, EFS and OS) a sensitivity analysis may be performed for the CR subjects who have undergone stem-cell transplantation by censoring them at their last disease assessment prior to the transplant.

Transplant-naïve subjects to proceed to further autologous/allogeneic stem cell transplantation will be measured by the - proportion of eligible subjects reaching autologous or allogeneic transplantation. The eligible subjects are defined as the transplant-naïve subjects who have achieved CR and then proceed to further autologous/allogeneic stem cell transplantation.

6.5 Efficacy Subgroup Analyses

The subgroup analyses will be performed for ORR (overall response rate) for the full analysis set defined in Section 5.2. Subgroup analyses will be detailed in the SPP.

7.0 Safety Analyses

7.1 General Considerations

A safety analysis will be performed for all subjects who take at least one dose of study drug (venetoclax or ibrutinib). Safety evaluations include but are not limited to capturing adverse events (AEs), clinical laboratory testing (hematology and chemistry), vital sign measurements, electrocardiogram (ECG) testing (as clinically indicated), as a measure of safety and tolerability for the entire study duration. Throughout the study, AEs will be evaluated by the NCI-CTCAE v. 5.0 and summarized. Laboratory test results, vital signs and ECG assessments will be analyzed as appropriate.

7.2 Analysis of Adverse Events

All summaries/analyses involving AEs will include treatment-emergent adverse events (TEAE) only, unless otherwise specified. TEAE are defined as any event with onset after the first dose of study drug and no more than 30 days after the last dose of study drug. Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent, unless the study drug start time and the AE start time are collected and the AE start time is prior to the study drug start time. If an incomplete onset date was collected for an AE, the AE will be assumed to be treatment-emergent unless there is evidence that confirms that the AE was not treatment-emergent (e.g., the AE end date was prior to the date of the first dose of study drug).

For summaries of AEs related (reasonable possibility) to study drug, at each level of summation (overall, SOC, and PT) each subject is counted only once. If a subject has an AE with unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship of "no reasonable possibility" present. The only exception is if the subject has another occurrence of the same AE with the relationship of "reasonable possibility." In this case, the subject will be counted under the reasonable possibility category.

Adverse event data will be summarized and presented using primary MedDRA system organ classes (SOCs) and preferred terms (PTs) according to the MedDRA coding dictionary version 21.0 or higher.

The number and percentage of subjects experiencing treatment-emergent adverse events will be summarized for the following adverse event summaries:

- Any treatment-emergent adverse event.
- Any treatment-emergent adverse event with NCI toxicity (CTCAE version 5.0) grade 3, 4, or 5 adverse events.
- Any treatment-emergent adverse event with NCI toxicity (CTCAE version 5.0) grade 3 or 4 adverse event.
- Adverse events broken down by NCI toxicity grade (Severity).
- Any treatment-emergent serious adverse event.
- Any treatment-emergent adverse event leading to discontinuation of venetoclax.
- Any treatment-emergent adverse event leading to venetoclax interruption.
- Any treatment-emergent adverse event leading to venetoclax reduction.
- Any treatment-emergent adverse event that is rated at least possibly related to venetoclax by the investigator (Reasonable Possibility Related)
- Any treatment-emergent adverse event leading to discontinuation of ibrutinib.
- Any treatment-emergent adverse event leading to ibrutinib interruption.
- Any treatment-emergent adverse event leading to ibrutinib reduction.
- Any treatment-emergent adverse event that is rated at least possibly related to ibrutinib by the investigator (Reasonable Possibility Related)
- Any treatment-emergent adverse event leading to death.

In addition, selected adverse events will be summarized. The list of such events is shown in Table 2.

Table 2. Selected Adverse Events

Risk	Search Criteria
Tumor Lysis Syndrome (TLS)	SMQ – "Tumour lysis syndrome" (Narrow-scope)
Neutropenia	PT terms – "Neutropenia," "Neutrophil count decreased," "Febrile neutropenia," "Agranulocytosis," "Neutropenic infection," and "Neutropenic sepsis"
Serious Infection, Including Opportunistic Infections	SAEs in the SOC of "Infections and Infestations"
Second Primary Malignancy	SMQ – "Malignant tumours" (Narrow) and "Myelodysplastic syndromes" (Narrow)
Lymphopenia	PT terms – "Lymphopenia" and "Lymphocyte count decreased"
Anemia	PT terms – "Anaemia" and "Haemoglobin decreased"
Thrombocytopenia	PT terms – "Thrombocytopenia" and "Platelet count decreased"
Drug Induced Liver Injury (DILI)	SMQ – "Drug related hepatic disorders – comprehensive search"
Medication Error	SMQ – "Medication error" (broad)
Hemorrhage	SMQ – "Haemorrhagic disorders" (narrow)

7.3 Analysis of Laboratory Data

Definition of Baseline

The baseline value for laboratory variables will be defined as:

- For subject hospitalized for TLS prophylaxis, the baseline value will be the lab value taken at hospital admission prior to the subject receiving hydration for TLS prophylaxis (the day prior to the first dose of study drug).
- For subject not hospitalized for TLS prophylaxis, the baseline value will be the 0 hour lab draw prior to the first dose of study drug on Day 1.

Laboratory Assessments for TLS

To determine if a subject's laboratory values qualify for TLS, the Howard criteria will be assessed. The Howard definition for laboratory TLS requires ≥ 2 of the metabolic abnormalities specified in Table 3 post-baseline and within 24 hours of each other.

Table 3. Laboratory Criteria for TLS

Element	Value
Uric Acid	> 476 µmol/L (8 mg/dL)
Potassium	> 6.0 mmol/L (6 mEq/L)
Inorganic Phosphorus	> 1.5 mmol/L (4.5 mg/dL)
Calcium	< 1.75 mmol/L (7 mg/dL)

The following summaries of laboratory criteria will be provided:

- Number and percentage of subjects meeting the definition of laboratory TLS (at least two values meeting the criteria in Table 4, occurring within 24 hours of each other).
- Listing of all values for these four analytes for each subject meeting the definition of laboratory TLS at least once during treatment.

Assessment of Drug-Induced Liver Injury

The number and percentage of subjects in each treatment group who have at least one observed post-baseline value meeting the following criteria will be tabulated: ALT > 3 × ULN, AST > 3 × ULN, AST or ALT > 3 × ULN, total bilirubin value > 2 × ULN, ALT or AST > 3 × ULN and total bilirubin > 2 × ULN within 24 hours. A listing of all observed ALT, AST, and total bilirubin values will be generated for subjects with an observed value meeting any of these criteria.

Analysis of Shift from Baseline in Clinical Laboratory Data

For shifts relative to National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI CTCAE version 5.0), baseline and post-baseline laboratory observations (maximum and final) will be categorized as grade 0, grade 1, grade 2, grade 3, or grade 4 according to NCI CTCAE grade version 5.0. For laboratory tests for which a normal range limit is one end of the grade 1 range then values that are either within the normal range or outside it in opposite direction will be classified as grade 0 values.

The baseline and final grades will be defined respectively as the grade of the last measurement collected prior to the first dose of study drug unless specified differently, and as the last post-baseline measurement collected no more than 30 days after the last dose of study drug. The maximum NCI toxicity grade value is the value with highest NCI toxicity grade collected after the first dose of study drug and within 30 days following the last dose of study drug. In cases where multiple values are collected on the same day, the maximum grade value will be selected as the value for that day.

For each variable, shift tables will be generated that cross tabulate the number of subjects with baseline values of grade 0, grade 1, grade 2, grade 3, or grade 4 versus maximum or final observations of grade 0, grade 1, grade 2, grade 3, or grade 4.

Detailed listings of data for subjects experiencing NCI CTCAE grade 3 to 4 blood chemistry and hematology values will be provided. All measurements collected, regardless of the number of days after the last dose of study drug, will be included in these listings.

7.4 Analysis of Vital Signs

For selected vital signs variables, a listing of all observations collected (at both scheduled and unscheduled visits) will be generated for subjects that had at least one post-baseline observation meeting pre-defined criteria for potentially clinically significant values. The number and percentage of subjects who have at least one post-baseline observation

meeting the pre-defined criteria for potentially clinically significant values will be provided for each variable.

Pre-defined criteria for potentially clinically significant vital signs values are given in Table 4:

Table 4. Criteria for Potentially Clinically Significant Vital Signs – Vital Signs Variables

Vital Signs Variables	Criterion	Definition of Potentially Clinically Significant
Systolic blood pressure	High	Value ≥ 160 mmHg
Diastolic blood pressure	High	Value ≥ 100 mmHg
Heart rate	Low	Value < 50 bpm
	High	Value ≥ 120 bpm
Temperature	Low	Value < 36°C
	High	Value ≥ 38.5°C

7.5 Analysis of ECG Parameters

Each ECG will be evaluated to determine if any findings outside normal physiological variation are clinically significant.

The categories for ECG readings are:

- Normal ECG
- Abnormal ECG – Not clinically significant (NCS)
- Abnormal ECG – Clinically significant (CS)
- Unable to evaluate

ECG test will be required only for the screening assessment (or if clinically indicated). Due to limited collection of the data, abnormal ECG values will be presented in a data listing.

8.0 Summary of Changes

8.1 Summary of Changes Between the Previous Version and the Current Version

- Response criteria changed to clinical response criteria proposed by the international T-PLL study group as per protocol amendment 1.
- nPR deleted as per protocol amendment 1.
- SAP-s changed to SPP as per the new SAP process.

8.2 Summary of Changes in Previous Version

Version	Date	Summary
1.0	05 Nov 2018	Original version
2.0	17 Jul 2019	Amendment 1

9.0 Appendix

List of Abbreviations

ABT-199	Study Drug Compound, "Venetoclax" GDC-0199
AE	Adverse Event
CR	Complete Remission
CRi	Complete remission with incomplete bone marrow recovery
CTCAE	Common Terminology Criteria for Adverse Events
C_{trough}	Trough concentrations
DoR	Duration of Response
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
EFS	Event-Free Survival
IRT	Interactive Response Technology
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NCS	Not Clinically Significant
ORR	Overall Response Rate
OS	Overall Survival
PFS	Progression-Free Survival
PK	Pharmacokinetics
PR	Partial Remission
PT	Preferred Term
QA	Quality Assurance
QC	Quality Control
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SOC	System Organ Class
SPP	Statistical Programming Plan
TEAE	Treatment-emergent Adverse Events
T-PLL	T-cell prolymphocytic leukemia
TLS	Tumor Lysis Syndrome

10.0 References

1. Shan G, Wilding GE, Hutson AD. Optimal adaptive two-stage designs for early Phase II clinical trials. Stat Med. 2016;35(8):1257-66.