STATISTICAL ANALYSIS PLAN

VERSION 3.0

CLINICAL STUDY PROTOCOL CP-MGD006-01

PROTOCOL AMENDMENT 12 (18 AUGUST 2020)

A PHASE 1/2, FIRST-IN-HUMAN, DOSE ESCALATION STUDY OF MGD006, A CD123 X CD3 DUAL AFFINITY RE-TARGETING (DART) BI-SPECIFIC ANTIBODY-BASED MOLECULE, IN PATIENTS WITH RELAPSED OR REFRACTORY ACUTE MYELOID LEUKEMIA OR INTERMEDIATE-2/HIGH RISK MYELODYSPLASTIC SYNDROME

REVISION HISTORY

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse event of special interest
AML	acute myeloid leukemia
CR	complete remission (mCR, CRc, or CRm)
CR1	initial CR
CRc	cytogenetic complete remission
CRh	complete remission with partial hematologic recovery
CRi	complete remission with incomplete blood cell recovery
CRm	molecular complete remission
CRn	CR with incomplete neutrophil recovery
CRp	CR with incomplete platelet recovery
CRS	cytokine release syndrome
CTCAE	common terminology criteria for adverse events
DLT	dose-limiting toxicity
DOR	duration of response
DSM	Data Safety Monitor
EBV	Epstein-Barr Virus
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EFS	event-free survival
EOTV	end of treatment visit
ER	early relapse
FDA	Food and Drug Administration
HSCT	hematopoietic stem cell transplantation
IDMC	independent data monitoring committee
IRR	infusion related reaction
IWG	International Working Group
mCR	morphologic CR
MLFS	morphologic leukemia-free state
MRD	minimal residual disease
MTD	maximum tolerated dose
MTDS	maximum tolerated dose and schedule
OS	overall survival
PBMC	peripheral blood mononuclear cell
PD	progressive disease

PIF	primary induction failure
PK	pharmacokinetics
PR	partial remission
SAE	serious adverse event
SD	stable disease
SPP	statistical programming plan
TLFs	tables, listings and figures
TTR	time to response
WBC	white blood cell (count)

1 INTRODUCTION

Flotetuzumab (MGD006) is a CD123 x CD3 DART protein developed by MacroGenics, Inc., designed to target CD123-positive cells (including AML cells) for recognition and elimination by CD3-expressing T lymphocytes as effector cells.

This statistical analysis plan (SAP) provides a detailed and comprehensive description of the analysis of the study CP-MGD006-01 entitled "A Phase 1/2, First-in-Human, Dose Escalation Study of MGD006, a CD123 x CD3 Dual Affinity Re-Targeting (DART) Bi-Specific Antibody-Based Molecule, in Patients with Relapsed or Refractory Acute Myeloid Leukemia or Intermediate-2/High Risk Myelodysplastic Syndrome".

.

3 STUDY OBJECTIVES

3.1 Primary Objectives

To assess the anti-neoplastic activity of flotetuzumab in patients with PIF/ER AML, as determined by the proportion of patients who achieve CR/CRh.

3.2 Secondary Objective(s)

Collectively, the secondary objectives of this study are:

- Assessment of CR rate, CRh rate, overall complete response rate (CR, CRh, CRi [CRn, CRp], or MLFS), objective response rate (CR, CRh, CRi [CRn, CRp], MLFS, or PR), time to response and duration of response (DOR).
- To measure early mortality rates, overall survival (OS) and event-free survival (EFS).
- To determine the rate of eligible patients, per institution criteria, that transition to successful stem cell transplant after achieving complete response (CR, CRh, CRi [CRn, CRp], or MLFS).
- To assess rate of conversion to and maintenance of transfusion independence.
- To evaluate duration of initial hospitalization for flotetuzumab administration.
- To evaluate incidence and duration of hospitalizations subsequent to initial discharge.
- To monitor the safety and tolerability of flotetuzumab.
- To characterize the PK and immunogenicity of flotetuzumab.
- To determine safety and efficacy of tocilizumab in the treatment of IRR/CRS.

3.3 Exploratory Objectives

The exploratory objectives of the study are:

- To evaluate CD123 expression on blast cells
- To evaluate circulating cytokine levels at baseline and over time
- To evaluate circulating leukemic and normal cells at baseline and over time
- To evaluate circulating T lymphocyte populations and activation markers at baseline and over time
- To evaluate leukemic cells, leukemic stem cells, normal progenitor cells and T lymphocytes at baseline and over time in the bone marrow
- To evaluate molecular markers of minimal residual disease (MRD)
- To examine changes in T lymphocyte repertoire

- To evaluate the correlation between cytogenetic abnormalities with responses to flotetuzumab immunotherapy.
- To study adaptive immune changes during flotetuzumab treatment.
- To study the tumor microenvironment immune contexture.
- **Ruxolitinib Cohort Objectives:**
    - To characterize the onset, duration, and severity of IRR/CRS on an exploratory, pilot basis in flotetuzumab-treated patients receiving ruxolitinib and compare that to historical experience in patients not receiving ruxolitinib.
    - To measure and compare cytokine profile and T-lymphocyte populations in patients receiving the combination of ruxolitinib and flotetuzumab vs flotetuzumab alone.
    - To determine the safety and tolerability of the combination of ruxolitinib and flotetuzumab.

4 STUDY DESIGN AND PLAN

4.1 Overall Study Design and Plan

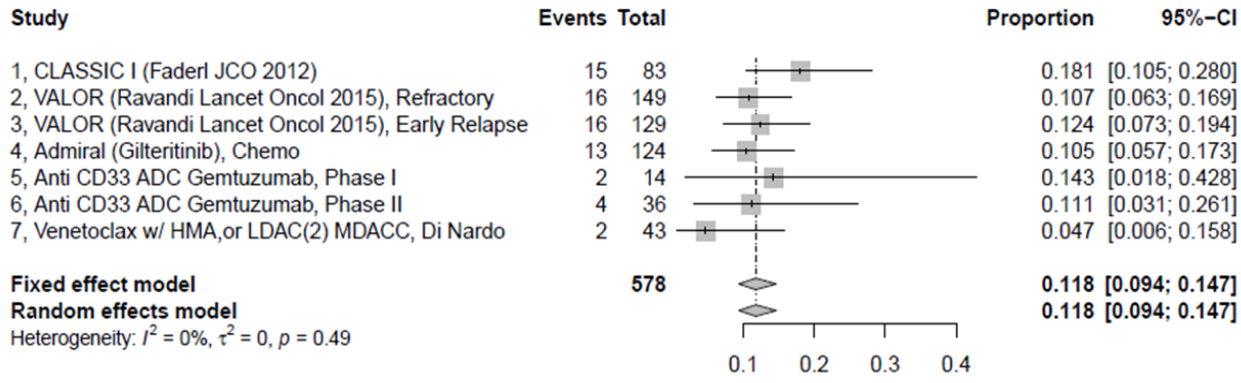
This is an open-label, multi-dose, single-arm, multi-center, Phase 1/2, dose-escalation and expansion study. As of Protocol Amendment 11 the trial design was updated to a single arm pivotal study going forward to evaluate efficacy and safety of flotetuzumab for PIF/ER AML patients.

4.2 Sample Size Determination

The planned sample size for the pivotal portion of the study as of Protocol Amendment 11 is 170 patients, based on a three-stage design.

To determine the sample size under inclusion criteria, historical reference response rates to standard treatments were ascertained through a meta-analysis based on 5 published studies that included PIF/ER AML patients selected from 963 publications identified through database search (**1**; **2**; **3**; **6**; **7**). The response rate identified by the meta-analysis is based on a random intercept logistic regression model, yielding a combined CR/CRh rate for PIF/ ER AML of 11.8%, with a 95% confidence interval of 9.4% to 14.7% (**Figure 1**).

Figure 1 Forest Plot for Meta-analysis



The result obtained with the Bayesian method is consistent with the result of the random intercept logistic regression model (**Table 1**).

Table 1 Meta-analysis Using Bayesian Method

Assumption	Point Estimate	95% Credible Intervals (Highest Posterior Density)	95% Credible Intervals (Equal-tail)
Normal prior, normal (0, var=10000)	11.7%	(9.3%, 14.7%)	(9.2%, 14.6%)
Beta prior, beta (1, 1)	11.9%	(9.4%, 14.6%)	(9.4%, 14.7%)

The upper 95% CI of 14.7% was chosen as the rate for the null hypothesis (H0) and a target of 23% CR/CRh was established as a clinically meaningful endpoint in this patient population.

A three-stage design with two interim analyses was employed to test the null hypothesis (H0) of CR/CRh rate = 14.7% against the alternative hypothesis (H1) of CR/CRh rate = 23%. Demonstration of CR/CRh rate improvement (H1) by rejecting H0 at a 1-sided type I error of 0.025 with adequate power (80%) in a sequential design requires 170 patients. Specifically, the planned number of patients and the futility and efficacy boundaries at each stage are as follows:

a. **First interim analysis:** The first interim analysis will be performed after 40 patients have been enrolled and either have had their responses evaluated or were discontinued for any reason from study treatment prior to their first response assessments. This interim analysis is for futility only, that is, the study enrollment will continue if at least 5 (12.5%) out of 40 patients have response of CR/CRh.

b. **Second interim analysis:** If the enrollment continues after the first interim analysis, a second interim analysis will be performed after additional 60 patients have been enrolled (100 patients in total) and either have had their response evaluated or were discontinued for any reason from study treatment prior to their first response assessment. This interim analysis is for both futility and efficacy. Specifically,

    i. If the number of CR/CRh is less than 15 (15%) out of 100 patients, then the study is considered futile and enrollment will be stopped.

    ii. If the number of CR/CRh is 15 (15%) or greater out of 100 patients, then enrollment will continue.

iii. If the number of CR/CRh is 25 (25%) or greater out of 100 patients, then H0 is rejected and the study is considered positive. The enrollment will still continue.

c. **Final analysis:** If enrollment continues after the second interim, then the final analysis will be performed after additional 70 patients have been enrolled (170 patients in total) and either have had their response evaluated or were discontinued for any reason from study treatment prior to their first response assessments. If the number of CR/CRh is at least 35 (20.6%) out of the total of 170 patients, then H0 is rejected and the study is considered positive.

The above design yields a 1-sided alpha of 0.005 for the efficacy boundary at the second interim analysis and an overall 1-sided alpha of 0.024 with approximately 80% power at the final analysis (**Table 2**). The probability of early termination under H0 (futility) in the first two stages is 55%.

Table 2 Three Stage Design

H0 (Benchmark CR/CRh rate)	H1 (Target CR/CRh rate)	Stage 1		Stage 2			Efficacy interim analysis at the end of Stage 2		Stage 3			1-Sided alpha for rejecting H0	Power under H1
14.7%	23%	N1	Move to Stage 2 if # (%) of CR/CRh ≥	N2	N1+N2	Move to Stage 3 if # (%) of CR/CRh ≥	Trial success if # (%) of CR/CRh ≥	1-sided alpha for rejecting H0	N3	Total (N1+N2+N3)	Trial success if # (%) of CR/CRh ≥	0.024	80%
		40	5 (12.5%)	60	100	15 (15.0%)	25 (25.0%)	0.005	70	170	35(20.6%)		

5 STUDY POPULATIONS

Three populations will be used for analysis, the efficacy population, safety population and the evaluable population, as defined below:

- **Efficacy population** – All patients that have been enrolled under protocol amendment 11 or later and treated with flotetuzumab at the MTD of 500 ng/kg/day as a continuous 7-day per week IV infusion during Cycle 1, have received any portion of one dose of flotetuzumab, and met the definition of PIF/ER AML based on Amendment 11 inclusion criteria . This population will be used for the primary analysis of efficacy endpoints.
- **Safety population** – All patients who received any portion of any dose of flotetuzumab. AML and MDS will be analyzed as pooled and separate populations for safety and PK. The analysis of all safety endpoints will be based on the safety population.
- **Response evaluable population** – All patients in the efficacy population that have a baseline bone marrow assessment, have at least one post-infusion assessment of their disease status or were removed from the study for reason of documented evidence of disease progression, death, or treatment-related adverse event. This population will be used in the summary of response rate as a secondary analysis.

Patients who die before having a post-infusion disease assessment will be considered evaluable for OS and EFS analyses.

6 STUDY ENDPOINTS

6.1 Primary Efficacy Endpoint

The primary efficacy endpoint for the PIF/ER AML pivotal expansion cohort is the rate of CR/CRh as assessed by the study investigators.

6.2 Secondary Efficacy Endpoints

Key secondary endpoints are listed in **Table 3**.

Table 3 Key Secondary Endpoints

Name	Description
Duration of response (DOR)	Time of initial documentation of response to the time of disease relapse or death due to any cause, whichever occurs first.
Rate of post-baseline transfusion independence	Rate of no transfusion during any consecutive 56-day period on-study.
Rate of HSCT	Rate of successful HSCT through study treatment.
Incidence and duration of hospitalization	Duration of initial hospitalization as well as incidence, number and durations of subsequent hospitalizations following the initial discharge, number of days hospitalized, and reasons for hospitalization.
CR rate	Rate of complete remission.
CRh rate	Rate of complete remission with partial hematologic recovery.

The additional secondary endpoints are listed in **Table 4**.

Table 4 Additional Secondary Endpoints

Name	Description
Overall complete response rate	Rate of achievement of BM blast of < 5%, CR+CRh+CRi [CRn, CRp]+MLFS according to IWG AML response criteria.
Objective response rate	Rate of CR+CRh+CRi [CRn, CRp]+MLFS+PR according to IWG AML response criteria.
Overall survival (OS)	Time from first dose to death from any cause.
Event-free survival (EFS)	Time from first dose to the date of relapse after CR/CRh/CRi [CRn, CRp], or death from any cause, whichever occurs first.
Time to response (TTR)	Time from first dose to the first CR, CRh, CRi (CRn, CRp), or MLFS.
Mortality rate at 30, 60, 90 and 180 days	Rate of death from any cause within 30, 60, 90, 180 days of first study dose.
Six-month survival rate	Probability of survival at 6 months from first study dose.
One-year survival rate	Probability of survival at 1 year from first study dose.
Immunogenicity	Proportion of patients who become positive for anti-drug antibody production.
Pharmacokinetics	Evaluation of serum concentration of flotetuzumab and its relationships with efficacy, safety and correlative endpoints.

6.3 Safety Endpoint

Safety assessments include evaluation of type, frequency, severity, and relationship to flotetuzumab treatment of AEs, SAEs and AESIs. These events will be recorded by the Investigators in the electronic case report forms (eCRFs). Verbatim terms will be coded to lower-level terms in the Medical Dictionary for Regulatory Activities (MedDRA) and graded by the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 4.0, except for infusion related reaction/cytokine release syndrome events which are defined by the modified criteria proposed by Lee et al (**5**). The most current version of MedDRA will be used. Additionally, the safety evaluation includes clinical laboratory evaluations, death, physical examination, vital signs, ECOG performance status, electrocardiogram (ECG) and concomitant medications/surgeries/procedures/therapies.

7 STATISTICAL METHODOLOGY

7.1 General Considerations

Safety and efficacy summaries will be provided for each dose level cohort.

Unless otherwise specified, the following general considerations are applied in data analyses:

- All statistical tests will preserve a significance level of 0.05 for two-sided tests, unless otherwise specified.
- The baseline value is defined as the latest value at the time of or prior to the first dose of study treatment.
- Categorical data will be summarized by the number and percent of patients falling within each category.
- Continuous variables will be summarized by descriptive statistics, including mean, standard deviation, median, minimum, and maximum.
- Time-to-event endpoints will be summarized by the number and percent of the event, median time and corresponding 95% confidence interval (CI), and event free rate and corresponding 95% CI at the specified time points of interest.
- To summarize safety/efficacy data by visit, the post-baseline results will be summarized by the scheduled visit or the derived visit as appropriate.
- In general, the derived visit is the same as the scheduled visit in the protocol which is defined by study day; data that is closest to derived visit (before and after) will be used as the value for that visit. In case that two data collected at the equal distance to the derived visit, the later one will be used.
- Summaries in the Safety Population will be reported by dose cohort and total population; summaries in the Efficacy Population will be reported by PIF, ER and the combined PIF/ER total population.
- The impact of COVID-19 may be assessed and analyzed on a case by case basis. In particular, the following analyses may be performed if there are enough COVID-19 positive patients to render the analyses meaningful:
    - Number of patients discontinued study treatment and/or discontinued from study due to COVID-19, if such reason is collected
    - Summary of tumor response, PFS and OS
    - Summary of study drug interruption, delay or withdraw due to COVID-19, if such reason is collected
    - Listing of protocol deviation due to COVID-19 separately

- All data summaries and tabulations will be conducted using SAS® software Version 9.3 or higher.

7.2 Missing Data

Data that are reported as missing will be treated as missing in all data summaries. Imputation rules for partially recorded dates, in cases where complete dates were required to carry out an analysis, will be described in the Statistical Programming Plan (SPP). In descriptive summaries for safety, observations that appear spurious (extreme values relative to the majority of the data) will not be altered or removed from the summary.

8 PATIENT DISPOSITION AND BASELINE CHARACTERISTICS

8.1 Patient Disposition

For patient disposition, the number and percentage of patients who reach various study milestones are summarized: All screened patients are broken down by screen failures (with reasons if collected) and enrolled. Then the category of enrolled patients is broken down by "never treated" (with reasons if collected) and "treated". The category of treated patients will further be broken down by "treatment ongoing" and "treatment discontinued" (with reasons for discontinuation, which also include protocol-defined treatment completion, if any).

A summary of patient disposition will be presented for both safety and efficacy population.

8.2 Patient Demographics and Baseline Characteristics

8.2.1 Demographics

Age (year), height (cm), weight (kg), body mass index (kg/m^2), and other continuous baseline characteristics will be summarized using descriptive statistics (N, mean, standard deviation, median, minimum, maximum), while age group (< 70 years, ≥ 70 years), gender, ethnicity, race, geographic region and other categorical variables will be provided using frequency tabulations (count, percent) for both safety and efficacy population.

Age or year of birth will be recorded on the eCRF. Where age is not recorded, age will be calculated as follows: Age = year of informed consent– year of birth + 1.

Body mass index (BMI) will be calculated as follows: BMI (kg/m^2) = weight in kg/height in m^2.

8.2.2 Baseline Disease Characteristics

Baseline disease characteristics will be summarized using descriptive statistics.

1. Eastern Cooperative Oncology Group (ECOG) performance status;
2. Prior AML therapies;
3. Primary induction failure (yes versus no);
4. Early relapse (yes versus no), if yes, duration of initial CR (CR1);
5. Relapse (yes versus no);
6. Initial AML subclassification;
7. Genetic and cytogenetics abnormalities at baseline;
8. Prior chemotherapy/immunotherapy (not related to study indication, for patients with treatment-related secondary AML) (yes versus no);

9. Prior radiotherapy therapies (Not Related to Study Indication, for patients with treatment-related secondary AML) (yes versus no);
10. Prior surgery/procedure (Not Related to Study Indication) (yes versus no);
11. Bone marrow blasts (%) and category (< 20%; 20% to < 30%; 30% to < 50%; ≥ 50%) from the screening bone marrow aspirate sample;
12. Peripheral blood blast (%);
13. Baseline values of the following lab parameters: hemoglobin, platelet count, absolute neutrophil count (ANC), white blood cells (WBC);
14. Number of red blood cell (RBC) transfusions and number of units transfused, and number of platelet transfusions and number of units transfused within 28 days prior to first dose;
15. The medical and surgical history will be summarized by the Medical Dictionary for Regulatory Activities (MedDRA) latest version by system organ class (SOC) and preferred term (PT).

8.3 Prior/Concomitant/Follow-up Medications/Procedures/Therapies

Prior medications/procedures/transfusions are defined as medications/procedures/transfusions that were started prior to first dose or have start dates missing.

Concomitant medications/procedures/transfusions are defined as non-study medications/procedures/transfusions that are started on or after first dose but before EOTV date or started before the start of the study treatment and ended or remain ongoing during the study treatment. Medication/procedure/transfusion with missing start date but end date on or after first dose date and before EOTV date or with both missing start date and end date is also considered as concomitant medication/procedure/transfusion.

Follow-up medications/transfusions are defined as non-study medications/transfusions that are started after EOTV date.

8.3.1 Prior/Concomitant/Follow-up Medications

All prior, concomitant and follow-up medications will be summarized in frequency tabulations (patient counts and percentages) and by World Health Organization ATC first level and PT. All tables described in this section will be produced for both safety and efficacy population.

8.3.2 Prior/Concomitant/Follow-up Surgeries/Procedures

All prior, concomitant and follow-up procedures/surgeries will be summarized in frequency tabulations (patient counts and percentages) and by SOC and PT. All tables described in this section will be produced for both safety and efficacy population.

8.3.3 Prior/Follow-up Anti-Cancer Therapies

Anti-cancer therapies include radiotherapy, induction/consolidation/ maintenance/salvage therapies, chemotherapy and immunotherapy. A listing will be provided for anti-cancer therapies.

8.3.4 Prior/Concomitant/Follow-up Transfusions

The number of patients receiving any transfusion product, and the number of patients receiving any transfusion type (red blood cells (PRBCs), whole blood, platelets, plasma, other) will be summarized by dose cohort groups. Prior /concomitant/follow-up transfusions will be summarized in separate tables for both safety and efficacy population.

Transfusion type, number of units, reasons for taking transfusion, and dates of transfusions will be listed in a listing.

9 STUDY DRUG EXPOSURES

All analyses of treatment exposure will be conducted using the safety and efficacy population.

The summary of study drug exposure will include descriptive statistics as well as frequency counts for the number of doses or cycles received, the total dose actually administered as well as the total dose intended, and the dose intensity, which is calculated as percentage of total dose actually administrated divided by total dose intended during whole treatment period up to the last dose of study drug. Dose intensity by cycle will be summarized.

Duration of study treatment exposure in days will be calculated as: date of EOTV – first dose date + 1 for patients who have discontinued treatment; date of data cutoff – first dose date + 1 for patients whose treatment is ongoing.

10 PROTOCOL DEVIATIONS

Critical protocol deviations will be identified prior to database lock for final analysis and will be listed and summarized.

11 EFFICACY ANALYSES

Disease activity will be assessed by the study investigators by using CBC and peripheral blood cell morphological examination, examination of bone marrow aspiration (or biopsy if required), and physical examination. Modifications to the Revised Recommendations of the International Working Group for Diagnosis (IWG), Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia will be used (**Appendix 1**). Response assessment will be based on bone marrow aspirate/biopsy and the best CBC up to 14 days post bone marrow aspirate/biopsy. Responses will be categorized as:

- CR: mCR (morphologic CR], CRc (cytogenic CR], or CRm (molecular CR);
- CRh (CR with partial hematologic recovery);
- CRi (CR with incomplete blood count recovery): CRn (CR with incomplete neutrophil recovery) or CRp (CR with incomplete platelet recovery);
- MLFS (morphologic leukemia free state);
- PR (partial response);
- OB (other benefit);
- SD (stable disease);
- PD (progressive disease).

The number and percent of PIF/ER AML patients will be summarized by best overall response. Efficacy analyses will be based on the PIF/ER AML population . All efficacy results will be summarized by PIF, ER and the total combined PIF/ER populations.

11.1 Primary Efficacy Analysis

The primary efficacy endpoint for the AML expansion cohort is the CR/CRh rate, calculated as the proportion of patients in the efficacy population that has achieved a best response of CR (mCR, CRc, or CRm) or CRh at any point during treatment by investigator's assessment, according to IWG AML response criteria (**Appendix 1**). CR/CRh rate will be summarized for total combined PIF/ER patients as the primary analysis, as well as for PIF and ER patients, separately, and for PIF/ER patients in the response evaluable population as a sensitivity analysis. A two-sided exact 95% confidence interval will be calculated around response rates.

For the second interim analysis of 100 patients, as well as for the final analysis of 170 patients, all treated patients will be followed for at least 6 months or death from any cause, whichever comes first, prior to data cut off for the analysis. If the trial passes the pre-specified efficacy boundary (25% observed CR/CRh rate) at the second interim analysis, a two-sided exact 95% confidence interval described by Jennison and Turnbull (**44**) will be produced to account for the bias in the conventional exact confidence interval for fixed sample test.

11.2 Secondary Efficacy Analyses

Counts, percentages, and two-sided exact 95% confidence intervals will be used to describe categorical secondary variables. Kaplan-Meier (KM) methods will be used to estimate time-to-event endpoints, unless otherwise specified. All secondary efficacy endpoints will be summarized in both efficacy and response evaluable population by PIF, ER, and total.

11.2.1 Key Secondary Efficacy Analyses

The key secondary efficacy endpoints include CR rate, CRh rate, duration of response, post-baseline transfusion independence rate, HSCT rate and incidence and duration of hospitalization.

Duration of response (DOR) will be calculated from the time of initial documentation of response to the time of disease relapse or death due to any cause, whichever occurs first. Patients who are still in response at study completion without alternative therapies will be censored at the time of their last disease assessment.

Duration of response will be calculated for patients who achieve the first evidence of CR or CRh as primary analysis. A separate exploratory analysis will be conducted for patients who achieve the first evidence of complete response (CR, CRh, CRi [CRn, CRp], or MLFS).

DOR (months) = (date of event [disease relapse or death] or date of censoring –date of first evidence of response + 1) / (365.25/12)

Sensitivity analyses will be done for duration of response by using the additional censoring rules below:

- Censored at the date of HSCT.
- Censored at the date of other alternative therapies such as surgery, procedure, radiation and subsequent antileukemic therapy, etc.
- Patients with two or more consecutive missing response assessments prior to event will be censored at the last date of response assessment before the missing assessments.

The duration of response will be analyzed using the KM method. The 25^{th} percentile, median and 75^{th} percentile time (including two-sided 95% CI) will be summarized.

Post-baseline transfusion independence is defined as no transfusion during any consecutive 56-day period on-study. Post-baseline transfusion independence will be summarized for RBC and platelet transfusion separately and in total. Transfusion dependence at baseline is defined as any red blood cell or platelet transfusions during the 28 days prior to the first dose of study drug. The rate of conversion from transfusion dependence to transfusion independence will be calculated. The rate of patients who are transfusion independent at baseline and remain independent during any 56-day post-baseline period will also be calculated.

HSCT rate is defined as rate of successful HSCT through study treatment. Patients who received other therapies (except for conditioning therapy) for AML after discontinuation of study drug and later undergo a subsequent HSCT will not be included. Number and percentage of patients with HSCT will be summarized.

Incidence and Duration of Hospitalization, summaries will be provided for the efficacy population, as well as independently for patients who achieved CR or CRh. Summaries will include duration of initial hospitalization as well as incidence, number and duration of subsequent hospitalizations following the initial discharge, number of days hospitalized, and the reasons for hospitalization.

11.2.2 Additional Secondary Efficacy Analyses

Additional secondary efficacy endpoints include overall complete response rate, objective response rate, subgroup analyses of CR/CRh, CR, CRh and overall complete response rates, overall survival, event-free survival, time to response, mortality rate at 30, 60, 90, and 180 days, six-month and one-year survival.

Overall complete response rate is defined as the rate of CR+CRh+CRi [CRn, CRp]+MLFS according to IWG AML response criteria (**Appendix 1**), and includes any response which achieves < 5% BM blasts by morphology.

Objective response rate is defined as the rate of CR+CRh+CRi [CRn, CRp]+MLFS+PR according to IWG AML response criteria (**Appendix 1**).

Subgroup analyses of CR/CRh, CR, CRh, and overall complete response rates will be performed by baseline disease characteristic (e.g., age group [< 70, ≥ 70], gender, cytogenetic risk status [high, intermediate, low], prior line of therapy [1, 2, 3], TP53 gene mutations)

Overall survival (OS) is defined as the time between the date of the first dose of study drug and death from any cause. Patients who are alive or whose death status is unknown will be censored at the last date known alive on or before data cutoff date. The last known-to-be-alive date is the last non-imputed date of any subject record in the study database. This date may be the last visit date or last contact date that the subject is known to be alive.

OS (months) = (date of death or last known alive – first dose date +1) / (365.25/12)

Event-free survival (EFS) is defined as the time from the first dose of study drug until date of evidence of relapse from CR, CRh, or CRi (CRn, CRp), or death from any cause, whichever occurs first. Patients who are not known to have any of these events at study completion will be censored at the date of the last disease assessment.

EFS (months) = (date of event [relapse or death] or date of censoring – first dose date + 1) / (365.25/12)

Sensitivity analysis of OS and EFS: patients will be censored for alternative therapies or procedures such as surgery, HSCT, radiation and subsequent antileukemic therapy, etc. If a patient begins a subsequent antileukemic therapy prior to event, the patient will be censored at the date of last assessment before starting the new anti-cancer therapy. For EFS, patients with two or more consecutive missing response assessments prior to event will be censored at the last date of response assessment before the missing assessments.

Time to response (TTR) is defined as time from first dose of study drug to first CR, CRh, CRi (CRn, CRp), or MLFS and will be summarized with descriptive statistics. Time to CR/CRh and time to best overall response will be calculated.

Mortality rate is defined as rate of death from any cause within 30, 60, 90, or 180 days of first dose of study drug. Number and percentage of deaths at 30, 60, 90, or 180 days will be summarized.

Six-month survival rate is defined as the probability of survival at 6 months from first dose of study drug.

One-year survival rate is defined as the probability of survival at 1 year from first dose of study drug.

Six-month and one-year survival rates will be evaluated by KM estimate with 95% CI.

11.2.3 Other Secondary Analyses

Immunogenicity

Anti-drug antibody production will be monitored using a bridging ELISA method. The proportion of patients who become positive in this assay will be reported.

Pharmacokinetics

Serum concentrations of flotetuzumab will be monitored using an electrochemiluminescence-based sandwich assay. PK analyses will be performed using industry standard software. Population PK modeling will be performed, and an appropriate model and model parameters will be described. This analysis will be performed by an outside consultant.

Safety and Efficacy with Use of Tocilizumab

The number of patients who received tocilizumab in treatment of infusion related reaction (IRR) and or cytokine release syndrome (CRS) will be reported. The incidence, severity, and duration of IRR/CRS will be summarized and will be broken down by patients who received or not received tocilizumab.

12 SAFETY ANALYSES

12.1 Treatment Emergent Adverse Events

Only treatment-emergent AEs, defined as any event that is temporally associated with administration of study product, will be summarized in tables. Events prior to treatment (e.g., due to study-related procedure) will be listed in an appendix to the final study report. DLTs will be listed.

The following tables of AE data will be created to summarize the number and percent of patients who experience at least one event of each of the following types:

- All AEs
- Drug-related AEs
- AEs by CTCAE Grade
- Drug related AEs by CTCAE Grade
- AEs with CTCAE Grade severity ≥ Grade 3
- Drug-related AEs with CTCAE Grade severity ≥ Grade 3
- SAEs (this may be a listing if there are few events)
- Drug-related SAEs
- Fatal AEs (this may be a listing if there are few events)
- AEs that result in study discontinuation
- AEs that lead to withdrawal of study drug
- AEs categorized as AESI and/or IREs

Tabulated data will display the number and percent of patients that experience individual events by System Organ Class (SOC) and Preferred Term (PT). Events will be displayed alphabetically for SOC and in descending order of overall PT incidence within each SOC.

An overall summary of AEs will display the number and percent of patients who experience at least one event of each of the following types:

- Any AE
- Any drug-related AE
- Any AE with CTCAE Grade severity ≥ Grade 3
- Any drug-related AE with CTCAE Grade severity ≥ Grade 3
- Any SAE
- Any drug-related SAE

- Any AE that results in study or study drug discontinuation
- Any fatal AE
- Any AESI
- Any IREs

The following AESI will be summarized:

- All cytokine release syndrome or infusion related reaction events
- Grade 2 or greater immune-related AEs including, but not limited to, events of pneumonitis, colitis, autoimmune hepatitis, arthritis, glomerulonephritis, myocarditis and cardiomyopathy, hypophysitis, or thyroiditis
- Tumor lysis syndrome (TLS)
- Neutropenic sepsis/fungemia
- Grade 2 or greater neurological events.
- Capillary leak syndrome
- Epstein-Barr Virus (EBV) reactivation

The following summaries will be provided for AESI:

- All AESI
- Study drug related AESI
- Serious AESI
- Study drug related serious AESI
- AESI with CTCAE grade of Grade ≥ 3
- Study drug related AEs with CTCAE severity Grade ≥ 3

12.2 Clinical Laboratory Evaluations

Summaries of laboratory values will display descriptive statistics for numerically quantified labs. Summaries will be grouped by lab panel (e.g., hematology, blood chemistry, and urinalysis) and will be displayed by visit for each lab parameter. Graphs of mean values over time or individual values at each time point may be used.

Shift tables will be used to display the percent of patients who have a shift in their lab values from normal at baseline to each post-baseline visit by CTCAE v4.0 severity Grade.

Listings of all clinical laboratory data with abnormal flags will be provided by patients and tests.

12.3 Physical Examination

Physical examination will be performed at screening, at the end of each treatment cycle, and at the EOTV. Descriptive statistics for height (screening only) and weight, and examination of skin, HEENT (head, eyes, ears, nose, and throat), lymph nodes, heart, chest, lungs, abdomen, extremities, and neurologic system, both observed values and changes from baseline (weight) will be summarized.

12.4 Vital Sign Measurements and Performance Status

Evaluation of vital signs includes include blood pressure, heart rate, respirations, pulse oximetry, and temperature at scheduled visits.

Vital sign measurements and Eastern Cooperative Oncology Group (ECOG) will be listed for each patient and by scheduled visit. Descriptive statistics for vital signs and ECOG, both observed values and changes from baseline, will be summarized.

12.5 Electrocardiograms

Twelve-lead ECGs will be obtained according to the schedules outlined in the latest version of the protocol in order to evaluate the potential cardiac effects of study drug, including QT interval prolongation. Recorded values of ECG parameters and change from baseline values will be summarized at each scheduled visit.

In addition, at baseline and maximum post-baseline visits, the proportion of patients having absolute QTc Fridericia value (QTcF) and QTcF interval (ΔQTcF) of the following categories will be presented:

- ≤ 450 ms
- > 450 ms to ≤ 480 ms
- > 480 to ≤ 500 ms
- > 500 ms

At maximum post-baseline visits, the proportion of patients who have an increase from baseline in ΔQTcF of the following categories will be presented:

- ≤ 30 ms
- > 30 to ≤ 60 ms
- > 60 ms

Central interpretation of ECGs will be used for data analysis purposes and will be summarized by presenting the shift from baseline to worst by cycle. The overall ECG interpretation will be displayed in cross–tabulations by dose cohort.

12.6 Left Ventricular Ejection Fraction

If the patient has a history or signs and symptoms of cardiac disease, an echocardiogram (or multiple-gated acquisition scan [MUGA] if indicated) will be obtained during the screening period, if one was not performed within 30 days before the start of treatment (Cycle 1 Day 1). An echocardiogram (or MUGA if indicated) may also be conducted during the study if clinically indicated. A listing will be provided at patient level.

12.7 Pulmonary Function Testing

PFT, including DLCO and FEV1, will be conducted during the screening period if testing was not performed within 30 days before the start of treatment (Cycle 1 Day 1). PFT may also be conducted during the study if clinically indicated. listings will be provided at patient level.

- During the COVID-19 pandemic, if PFT equipment is not available, a correlative stair-climbing test may be conducted instead, together with lung CT and echocardiogram irrespective of cardiac history. The patient must pass the stair-climbing test, with O_2 saturation $\geq$ 95%, lung CT with normal appearing lung parenchyma, and normal echocardiogram

.

13 INTERIM ANALYSIS

13.1 General Information

Following a three-stage design, the interim analyses will be performed after 40 and 100 patients have been enrolled and either have had their responses evaluated or discontinued for any reason from study treatment prior to their first response assessments. The first interim analysis is for futility only and the second interim analysis is for both futility and efficacy. See **Section 4.2** for more details.

13.2 Data Safety Monitor (DSM)

Prior to Protocol Amendment 13, an independent DSM (with disease-specific and/or immunotherapy expertise) was involved in all decisions regarding transition from the Single Patient Dose Escalation Segment to the Multi-patient Dose Escalation Segment and regarding the definition of MTDS and transition to the MTDS Cohort Expansion Segment of the study. In the Single Patient Dose Escalation Segment, safety data from all patients followed through Study Day 28 (Cycle 2 Day 1) was reviewed. Safety data from each dose escalation cohort in the Multi-patient Dose Escalation Segment was made available to the DSM after each cohort of patients had been followed for a minimum of 28 Study Days.

Safety evaluations was conducted on an ongoing basis during the Cohort Expansion Segment of the study. Review by Principal Investigators (or designee), the Medical Monitor and the DSM was conducted in regular intervals. The interval analysis utilized data from all cycles of treatment. If during the conduct of the interval cohort analysis, the Bayesian posterior probability is greater than 80% that the DLT event rate is more than 20%, enrollment would be paused and an ad hoc meeting of the Principal Investigators (or designee), the Medical Monitor, and the DSM would be held to consider dose reduction or study termination.

13.3 Independent Data Monitoring Committee

Starting with Protocol Amendment 13, an independent data monitoring committee (IDMC) is formed to conduct regular independent safety review. This replaces the independent DSM in place prior to this amendment.

At each IDMC safety review meeting, IDMC will receive and review aggregate safety data. The IDMC will determine whether the enrollment should be paused based on Bayesian posterior probability evaluation of DLT event rate. This posterior probability is calculated based on the following Bayesian binomial model: Suppose that the DLT rate p follows non-informative prior Beta(1,1), and that r out of n treated patients have experienced DLTs, then the posterior distribution for p is Beta(1+r, 1+n-r). Then, for each n in 8 patient segments, the number of patients with DLTs that would result in a pause of enrollment is the smallest integer r, satisfying $P(p>20\%|n, r)>80\%$, where $p\sim Beta(1+r, 1+n-r)$. **Table 5** below shows the number of patients with DLTs that would result in a pause of enrollment to assess toxicity.

Table 5 Stopping Bounds for Excess Toxicity Based on Bayesian Beta Binomial Model with Beta (1,1) Prior

Number of patients treated (n)	Number of patients with DLTs that would result in a pause of enrollment (r)
8	3
16	5
24	7
32	9
40	10
48	12
56	14
64	16
72	17
80	19
88	21
94	22

The IDMC will also hold a minimum of two safety and efficacy review meetings for two planned interim analyses (additional ad hoc meetings will be held as needed). At each meeting, in addition to above safety evaluation, IDMC will also receive and review aggregate efficacy data to determine whether enrollment should continue based on futility and efficacy boundaries described in **Section 4.2**. The details of IDMC formation and meeting conduct is described in a the **IDMC Charter**.

13.4 Statistical Approaches for Control of Alpha

The type I error rate will be controlled at the overall one-sided 0.025 level by using 3-stage design. The decision rules are as follows:

First stage: the study enrollment will continue if at least 5 (12.5%) out of 40 patients have response of CR/CRh.

Second stage: the study enrollment will continue if at least 15 (15%) out of 100 patients have response of CR/CRh.

Claim success if at least 25 (25%) out of 100 patients have response of CR/CRh at the second stage or at least 35 (20.6%) out of the total of 170 patients have response of CR/CRh at the final stage.

The 3-stage design yields a 1-sided alpha of 0.005 for the efficacy boundary at the second interim analysis and an overall 1-sided alpha of 0.024 with approximately 80% power at the final analysis. The probability of early termination under H0 (futility) in the first two stages is 55%.

14 LIST OF TABLES, LISTINGS, AND FIGURES

The list of tables, listings, and figures (TLFs) and associated shells planned for the clinical study report based on the analyses described in this SAP will be provided in a separate SPP, which will also include data reporting conventions and programming specifications for the development of these TLFs.

15 REFERENCES

1. **DiNardo CD, Rausch CR, Benton C, Kadia T, Jain N, Pemmaraju N, et al.** Clinical experience with the BCL2-inhibitor venetoclax in combination therapy for relapsed and refractory acute myeloid leukemia and related myeloid malignancies. Am J Hematol. 2018;93:401-407.

2. **Faderl S, Wetzler M, Rizzieri D, Schiller G, Jagasia M, Stuart R, et al**. Clofarabine plus cytarabine compared with cytarabine alone in older patients with relapsed or refractory acute myelogenous leukemia: results from the CLASSIC I Trial. J Clin Oncol. 2012;30(20):2492-9.

3. **FDA.** Xospata (gilteritinib) label. Accessed 01 June 2020. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/211349s001lbl.pdf

4. **Jennison C and Turnbull BW.** Group Sequential Methods with Applications to Clinical Trials. Boca Raton, Florida: Chapman & Hall/CRC. 2000;235-244.

5. **Lee DW, Gardner R, Porter DL, Louis CU, Ahmed N, Jensen M, et al**. Current concepts in the diagnosis and management of cytokine release syndrome. Blood 2014;10;124(2):188-95

6. **Medeiros BC, Tanaka TN, Balaian L, Bashey A, Guzdar A, Li H, et al.** A phase I/II trial of the combination of azacitidine and gemtuzumab ozogamicin for treatment of relapsed acute myeloid leukemia. Clin Lymphoma Myeloma Leuk. 2018;18(5):346-352.

7. **Ravandi F, Ritchie EK, Sayar H, Lancet JE, Craig MD, Vey N, et al.** Vosaroxin plus cytarabine versus placebo plus cytarabine in patients with first relapsed or refractory acute myeloid leukaemia (VALOR): a randomised, controlled, double-blind, multinational, phase 3 study. Lancet Oncol. 2015;16(9):1025-36.

Appendix 1 International Working Group (IWG) Response Criteria for Acute Myeloid Leukemia (AML) Modified for Assessment of AML in this Study

Response Criterion	Time of Assessment	Neutrophils* (µL)	Platelets* (µL)	Bone Marrow Blasts (%)	Other
Complete Remission (CR)	Day 25 + 3 days (each cycle)	≥ 1,000	≥ 100,000	< 5	Transfusion[#] Extramedullary disease (EMD)
• Morphologic CR (mCR)	Day 25 + 3 days (each cycle)	≥ 1,000	≥ 100,000	< 5	Transfusion[#] EMD
• Cytogenetic CR (CRc)	Day 25 + 3 days (each cycle)	≥ 1,000	≥ 100,000	< 5	Cytogenics – normal, EMD
• Molecular CR (CRm)	Day 25 + 3 days (each cycle)	≥ 1,000	≥ 100,000	< 5	Molecular – negative, EMD
CR with partial hematologic recovery (CRh)	Day 25 + 3 days (each cycle)	≥ 500	≥ 50,000	< 5	Transfusion[$] EMD
CR with incomplete blood count recovery (CRi)	Day 25 + 3 days (each cycle)	< 1,000	< 100,000	< 5	Transfusion[$] EMD
• CR with incomplete neutrophil recovery (CRn)	Day 25 + 3 days (each cycle)	< 1,000	≥ 100,000	< 5	Transfusion[$] EMD
• CR with incomplete platelet recovery (CRp)	Day 25 + 3 days (each cycle)	≥ 1,000	< 100,000	< 5	Transfusion[$] EMD
Morphologic leukemia-free state (MLFS)	Day 25 + 3 days (each cycle)	NA	NA	< 5	EMD
Partial remission (PR)	Day 25 + 3 days (each cycle)	> 1,000	> 100,000	> 50% decrease or decrease to 5-25	Blasts < 5% if Auer rod positive
Other benefit (OB)	Day 25 + 3 days (each cycle)	NA	NA	> 30% decrease	NA

Response Criterion	Time of Assessment	Neutrophils* (µL)	Platelets* (µL)	Bone Marrow Blasts (%)	Other
Stable disease (SD)	Day 25 + 3 days (each cycle)	NA	NA	Absence of CR, CRh, CRi, PR, MLFS, OB, and criteria for PD not met	NA
Progressive disease	NA	NA	NA	Evidence for an increase in bone marrow blast percentage and/or increase of absolute blast counts in the blood: • > 50% increase in marrow blasts over baseline (a minimum 15% point increase is required in cases with < 30% blasts at baseline; or persistent marrow blast percentage of > 70% over at least 3 mo; without at least a 100% improvement in ANC to an absolute level (>0.5 x 10^9/L [500/mL], and/or platelet count to >50 x 10^9/L [50 000/mL] non-transfused); or • > 50% increase in peripheral blasts (WBC x % blasts) to > 25 x 10^9/L (> 25 000/mL) (in the absence of differentiation syndrome)†; or • New extramedullary disease	**In general, at least 2 cycles of flotetuzumab should be administered to determine "progressive disease".** "Progressive disease" is usually accompanied by a decline in ANC and platelets, increased transfusion requirement, and decline in performance status or increase in symptoms
Recurrence/relapse	Relapse after CR is defined as a reappearance of leukemic blasts in the peripheral blood and ≥ 5% blasts in the bone marrow not attributable to any other cause (e.g., bone marrow regeneration after consolidation therapy)				

CR = complete remission (includes mCR, CRc, CRm).
CRi = includes neutrophil count ≥ 1000 or platelet count ≥ 100,000.
*Assessment is quantitated by BM biopsy and best CBC (up to 28 days post BM biopsy).
[#] No RBC and/or platelet transfusion within 5 days prior to response assessment.
[$] No platelet transfusion within 5 days prior to response assessment
† Certain targeted therapies, for example, those inhibiting mutant IDH proteins, may cause a differentiation syndrome, that is, a transient increase in the percentage of bone marrow blasts and an absolute increase in blood blasts; in the setting of therapy with such compounds, an increase in blasts may not necessarily indicate PD.