Protocol B1371019

A RANDOMIZED (1:1), DOUBLE-BLIND, MULTI-CENTER, PLACEBO CONTROLLED STUDY EVALUATING INTENSIVE CHEMOTHERAPY WITH OR WITHOUT GLASDEGIB (PF-04449913) OR AZACITIDINE (AZA) WITH OR WITHOUT GLASDEGIB IN PATIENTS WITH PREVIOUSLY UNTREATED ACUTE MYELOID LEUKEMIA

Statistical Analysis Plan
(SAP)

Version: 5.0

Date: 14-Apr-2020

TABLE OF CONTENTS

LIST OF TABLES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B1371019 is based on Protocol Amendment #5 dated 12Apr2019 and the Type B supplemental new drug application (sNDA) meeting with the Food and Drug Administration (FDA) on 23Mar2020.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable
2	The PRO sections were revised following comments from FDA. Minor clarifications were made in other sections.	Following protocol Amendment #2
3	The Pharmacogenomics Analysis Set was removed. Clarification regarding select endpoints were made (CRi or better, and CRh or better). Add analysis method for China subset patients.	Following Protocol Amendment #3 and satisfying CFDA requirement
4	Region (rest of world versus China) added as a stratification factor to the applicable efficacy analyses for the intensive study. Updated the definition of "on-treatment" for assessments and treatment-emergent adverse events.	Following Protocol Amendment #5
5	Add OS subgroup analysis.	Following FDA feedback

	Add sensitivity analyses for handling potential missing or unclean data due to COVID-19.	

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B1371019. This analysis plan is meant to supplement the study protocol. Any deviations from this analysis plan will be described in the Clinical Study Report.

The analyses of this study will include all data up to a data cutoff date which will be determined by the number of overall survival (OS) events. All summaries and analyses will include all data pertaining to visits/assessments performed up to and including the data cutoff date.

Due to cleaning activities, the final number of events might deviate from the planned number. The data cut-off date will not be adjusted retrospectively in this case.

2.1. Study Objectives

Primary Objective

- To demonstrate that glasdegib is superior to placebo in combination with azacitidine (non-intensive study) or cytarabine and daunorubicin (intensive study) in prolonging OS in subjects with untreated AML.

Secondary Objectives:

- To compare fatigue score post baseline as measured by MDASI AML/MDS in both treatment arms;
- To compare glasdegib versus placebo in combination with azacitidine (non-intensive study) or '7+3' (cytarabine and daunorubicin) in improving other clinical efficacy measures;
- To estimate the duration of response in both treatment arms;
- To estimate the time to response in both treatment arms in the Non-intensive study only;
- To compare Event-free Survival (EFS) in both treatment arms;
- To compare PRO measurements in both treatment arms;
- To evaluate the overall safety profile in both treatment arms;
- To evaluate laboratory abnormalities in both treatment arms;

- To characterize the PK of glasdegib;
- To characterize treatment effects on the QTc interval.

CCI

[redacted]

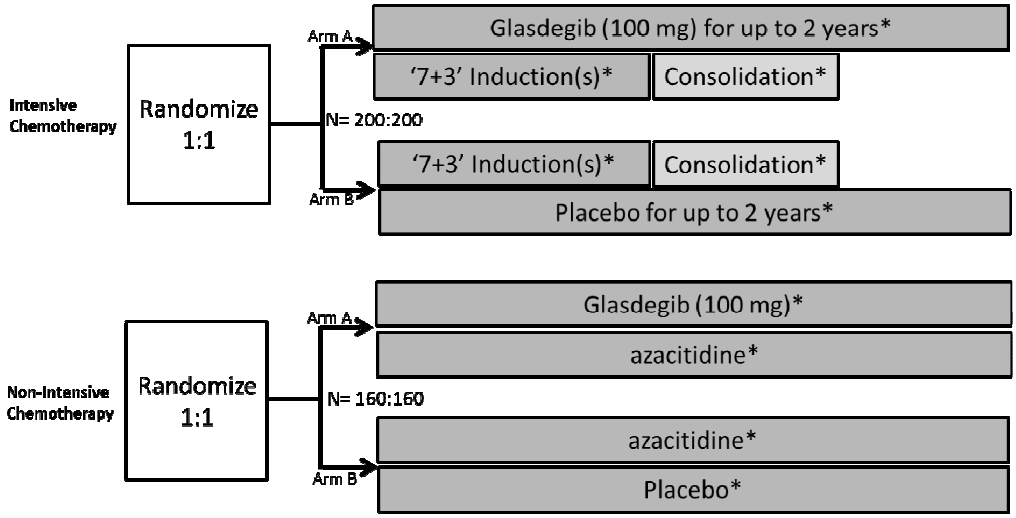
2.2. Study Design

Two separate registration trials conducted under one protocol number are proposed to adequately and independently evaluate the addition of glasdegib in intensive and non-intensive chemotherapy populations. Each study will have an experimental treatment arm and a placebo arm.

Study B1371019 (Figure 1) is a randomized (1:1), double-blind, multi-center, placebo controlled study of chemotherapy in combination with glasdegib versus chemotherapy in combination with placebo in adult patients with previously untreated AML (excluding APL with PML-RARA, AML with BCR-ABL1 and active central nervous system [CNS] leukemia) to support the proposed indication statements:

- Intensive Study: Glasdegib is being studied in combination with cytarabine and daunorubicin for the treatment of adult patients with untreated AML.
- Non-intensive Study: Glasdegib is being studied in combination with azacitidine for the treatment of adult patients with untreated AML who are not candidates for intensive induction chemotherapy.

Figure 1. Schematic of Study Design



* Protocol Section 5.5 describes dosing details for the intensive chemotherapy regimen(s) [including '7+3' or '5+2' option for Induction 2], non-intensive chemotherapy regimen, and experimental study drugs.

Assignment to the Intensive Study or the Non-Intensive Study will be made by the Investigator based on the 2017 ELN recommendations.

The treatment arm allocation for Arm A or Arm B within each study will be operated by the sponsor following the rules defined in Protocol Section 5.1.

Intensive Chemotherapy Study:

A total of 400 subjects eligible to receive intensive chemotherapy per investigator assessment will receive treatment. Subjects will be stratified by genetic risk (favorable vs intermediate vs adverse by ELN genetic risk categories) and age (≤60 years vs >60 years) at randomization.

A total of 267 death events would provide 90% power to detect an improvement in overall survival (translated in median OS [mOS] from 21 months to 31.5 months) assuming a composite mOS for the population as follows: young AML (aged ≤60 years) with a mOS of 23.7 months and elderly AML (aged >60 years) with a mOS of 15 months assuming a 70% young and 30% elderly AML split[12,13] and with an HR=0.67 using a 1-sided log-rank test at a significance level of 0.025 and a 3-look group-sequential design.

Two interim analyses will be conducted: an early efficacy and futility interim analysis of OS after 50% death events occur (no plan to stop for efficacy even if the efficacy boundary is crossed) and an efficacy and futility analysis of OS after 70% death events occur in the Intensive Chemotherapy population or upon completion of enrollment, whichever is later.

Non-Intensive Chemotherapy Study:

A total of 320 subjects who are not candidates to receive Intensive Chemotherapy will receive treatment. Subjects will be stratified by genetic risk (favorable vs intermediate vs adverse by ELN genetic risk categories) and age (<75 years vs ≥75 years) at randomization.

A total of 220 death events would provide 90% power to detect an improvement in overall survival (translated in mOS from 10.4 months[14] to 16.2 months) with an HR=0.64 using a 1-sided log-rank test at a significance level of 0.025 and a 2-look group-sequential design. One interim efficacy and futility analysis of OS would be planned after 60% death events occur in the non-intensive chemotherapy population or upon completion of enrollment of these subjects, whichever is later.

The interim and final analyses of the intensive chemotherapy and non-intensive chemotherapy populations will be conducted separately and independently of each other.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

Of note, the primary endpoint of OS is assessed based on all available data in the database and other secondary response endpoints are assessed by the investigator using local hematology, chemistry, bone marrow assessments, etc. Patients will complete PRO questionnaires in site tablets that will be collected by a central vendor. MRD results are determined by a central laboratory. Time to event and duration endpoints will be derived programmatically.

3.1. Primary Endpoint

OS is defined as the time from the date of randomization to the date of death due to any cause. Subjects last known to be alive will be censored at date of last contact.

3.2. Secondary Endpoints

3.2.1. Efficacy Endpoints

Secondary efficacy endpoints will be summarized based on data up to initiation of a new anti-cancer drug therapy, if applicable.

Fatigue

The response for the single item “Fatigue (tiredness)” (one of the 13 MDASI core cancer symptoms) will be analyzed as a key secondary efficacy endpoint. This specific item is measured at the patient’s WORST level in the last 24 hours and asks a patient to respond on a 0-10 numeric rating scale (NRS), where 0 = “not present” and 10 = “as bad as you can imagine”.

Disease specific efficacy endpoints

CR, $CR_{MRD\text{-}neg}$ (a subgroup of CR), CRi, Morphological Leukemia Free State, and Partial Remission, as defined according to the 2017 ELN recommendations, and CR with partial hematologic recovery (CRh) for the Non-intensive study only. Note that CR (including $CR_{MRD\text{-}neg}$) is a deeper response than CRh, and CRh is a deeper response than CRi. CR (including $CR_{MRD\text{-}neg}$) is always a deeper response than CRi.

Note that patients whose MRD status cannot be determined (eg, due to lack of proper identification of a baseline immunophenotype from aspirate or peripheral blood) will not be allowed to resume single-agent therapy post HSCT or post consolidation. An MRD status will not be determined and the best response possible will be CR with no MRD status associated with the response.

Following Consolidation or following HSCT, the MRD result must be confirmed by 2 consecutive central laboratory results. Once confirmed, this will be used as the MRD result.

Duration of Response

Intensive Chemotherapy Study

Duration of Response (DoRi) is only defined for subjects who have ever achieved CRi or better on study as the time from date of first achieving CRi or better to the date of relapse after CRi or better or death due to any cause. Subjects last known to be alive who are free from relapse after CRi or better are censored at the date of last disease assessment that verifies their status. Note that in this study CRi or better includes CR and CRi.

Non-Intensive Chemotherapy Study

DoRi is only defined for subjects who have ever achieved CRi or better on study as the time from date of first achieving CRi or better to the date of relapse after CRi or better or death due to any cause. We also define DoRh for subjects who have ever achieved CRh or better on study as the time from date of first achieving CRh or better to the date of relapse after CRh or better or death due to any cause. Note that in this study, CRi or better includes CR, CRh, and CRi, and CRh or better only includes CR and CRh.

Time to Response (Non-Intensive Study only)

Time to Response (TTRi) is only defined for subjects who have ever achieved CRi or better on study as the time from date of randomization to date of first achieving CRi or better. Similarly, TTRh is only defined for subjects who have ever achieved CRh or better on study as the time from date of randomization to date of first achieving CRh or better.

Event Free Survival

Intensive Chemotherapy Study

For the Intensive chemotherapy patients, Event Free Survival (EFS) is defined as the time from the date of randomization to the date of treatment failure (TF), relapse from CR, or death from any cause, whichever comes first. TF is defined as failure to achieve CR during the induction cycle (including the re-induction cycle if there is one) and the event date for TF is the day of randomization (ie, EFS of 1 day). Responders last known to be alive who are free from disease progression or relapse are censored at the date of last disease assessment that verifies their status.

CCI

Non-Intensive Chemotherapy Study

For Non-intensive chemotherapy patients, EFS is defined as the time from the date of randomization to the date of TF, relapse from CRh or better, or death from any cause, whichever comes first. TF is defined as failure to achieve CRh or better following up to 6 cycles of study treatment and the event date for TF is the day of randomization. Patients who discontinue either study treatment without achieving CRh or better prior to completing 6 cycles of study treatment are considered treatment failures. Responders last known to be alive who are free from disease progression or relapse are censored at the date of last disease assessment that verifies their status.

CCI

CCI

[redacted]

[redacted]

[redacted]

3.2.2. Safety Endpoints

AEs (graded by NCI CTCAE v.4.03 as provided by the investigator on the AE CRF page); laboratory abnormalities (graded by NCI CTCAE v.4.03 as programmatically derived based on laboratory values); vital signs (blood pressure, pulse rate) and body weight; electrocardiograms (ECGs); echocardiogram or multigated acquisition (MUGA) scan; ophthalmologic data.

3.2.3. PRO Endpoints

Patient reported outcomes will be assessed using the MD Anderson Symptom Inventory AML/MDS Module (MDASI-AML/MDS), EuroQol 5-Dimension questionnaire 5-Level version (EQ-5D-5L), Patient Global Impression of Symptoms (PGIS), and Patient Global Impression of Change (PGIC).

MDASI-AML/MDS

The MDASI-AML/MDS is a validated modulized patient reported outcome measure for AML and MDS. It consists of a 19-item core cancer module and a 4-item AML/MDS specific module. The 23 items are designed to measure 13 core cancer symptoms (pain, fatigue, nausea, disturbed sleep, distress, shortness of breath, problem remembering, lack of appetite, drowsiness, dry mouth, sadness, vomiting, and numbness), 4 AML/MDS-specific symptoms (malaise, diarrhea, muscle weakness, and skin problems), and 6 areas of interference (general activity, mood, work, walking, relations with other people, and enjoyment of life).

Patients are asked to rate the severity of symptoms and related interference at their WORST level in the last 24 hours by responding to each item on an 0-10 numeric rating scale (NRS), where 0 = “not present” or “did not interfere” and 10 = “as bad as you can imagine” or “interfered completely”.

For this study, the response for the single item “Fatigue” (one of the 13 MDASI core cancer symptoms) will be analyzed separately as the key secondary efficacy endpoint.

Note that the electronic system capturing PRO data is set up in such a way that no questions on a questionnaire can be skipped. Hence, the questionnaires are either 100% or 0% completed.

EQ-5D-5L

The EQ-5D-5L is a brief, self-administered, validated and reliable generic health status instrument developed by the EuroQoL Group. It consists of the EQ-5D descriptive system and a visual analogue scale (VAS), the EuroQoL visual analogue scale (EQ-VAS).

The EQ-5D descriptive system measures a patient's health state on 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The patient is asked to indicate his/her health state by rating each dimension on a five-level scale (1=no problem, 5=extreme problem). This rating resulted in a 5-digit number describing the respondent's health state. It should be noted that the numerals 1-5 have no arithmetic properties and should not be used as a cardinal score.

The EQ-5D Index scores are country specific and range generally from 0 to 1, with 0 representing the worst health state and 1 as perfect health. The index values, presented in country specific value sets, are a major feature of the EQ-5D instrument, facilitating the calculation of quality-adjusted life years (QALYs) that are used to inform economic evaluations of health care interventions.

The EuroQoL visual analogue scale (EQ VAS) records the respondent's self-rated health on a 20-cm vertical, visual analogue scale from 0 (worst imaginable health state) to 100 (best imaginable health state). This information could then be used as a quantitative measure of health as judged by the individual respondents.

PGIS

PGIS and PGIC are employed as anchors for defining a "responder threshold" for MDASI-AML/MDS item Fatigue. PGIS is generally the preferred anchor over PGIC because PGIC may be subject to recall bias. As recommended by the FDA, for this protocol, both anchor scales will be used to provide an accumulation of evidence to help interpret a clinically meaningful score change in MDASI-AML/MDS.

The PGIS is a single-item questionnaire designed to assess patient's overall impression of disease severity at a given point in time. It uses a 4-point Likert scale as follows: In the last 24 hours, my leukemia symptoms are: 1-"Absent (no symptoms)", 2-"mild", 3-" moderate", 4-"severe".

It will be used as an anchor for defining a "responder threshold", on MDASI-AML/MDS and can also be used to create severity categorization for this PRO to enhance interpretation.

PGIC

PGIC is a single-item questionnaire designed to assess the patient's overall sense of whether there has been a change since starting treatment as rated on a 7 point Likert scale anchored by (1) "very much improved" to (7) "very much worse", with (4)-"no change". The PGIC is a measure of "participant rating of global improvement and satisfaction with treatment".

This instrument is used to determine global improvement as assessed by the patient and as an anchor to define a responder definition for MDASI-AML/MDS and as a sensitivity analysis for defining a "clinical important difference" on this PRO.

3.2.4. Other Endpoints

- PK of glasdegib;
- QTc interval.

CCI

[redacted]

3.4. Baseline Variables

The date of first dose (start date) of study treatment is the earliest date of non-zero dosing of any study drug. The date of last dose of study treatment is the latest date of non-zero dosing of any study drug.

No windowing will be applied when defining baseline. Any deviations from the protocol specified window will be documented as protocol deviations.

For efficacy endpoints (except Fatigue) and baseline characteristics the last non-missing assessment prior to randomization will serve as the baseline assessment. For fatigue, similar to other PRO endpoints, the last non-missing measurement on the day of the first dose of study treatment will serve as baseline (note that PRO endpoints are collected prior to study drug dosing on the same day).

For safety (including Eastern Cooperative Oncology Group (ECOG) performance status) the last non-missing assessment performed on or prior to date of the first dose of study treatment (or prior to randomization for patients randomized but not dosed) will serve as the baseline assessment. If there are no observations meeting these criteria, then baseline is considered missing.

For PRO, the measurement, the day of the first dose of study treatment will be used as the baseline measurement. If there are no observations meeting this criteria, then baseline is considered missing.

Triplicate ECGs are collected; therefore the baseline for each ECG measurement is the average of the pre-dose measurements on the baseline day. Unscheduled assessments will not be included in the calculation of the average. The average of the replicate measurements will be determined after the derivation of the individual parameters at each timepoint.

3.4.1. Stratification

Randomization is stratified by the following factors:

Intensive Chemotherapy Study:

- genetic risk (favorable vs intermediate vs adverse by ELN genetic risk categories) and age (≤60 years vs >60 years).

Non-Intensive Chemotherapy Study:

- genetic risk (favorable vs intermediate vs adverse by ELN genetic risk categories) and age (<75 years vs ≥75 years).

The primary analysis will utilize strata as defined in the IVRS system. In addition, sensitivity analysis may be conducted with baseline age and ELN genetic risk category collected in the CRF.

3.5. Safety Endpoints

Safety endpoints will be summarized based on the on-treatment period unless otherwise specified.

On-treatment is defined as the time from the first dose of study treatment to last dose prior to HSCT + 28 days and if applicable, from first dose post-HSCT to last dose + 28 days or start of new anti-cancer drug therapy (minus 1 day), whichever occurs first. Note that conditioning regimens for HSCT (regardless of whether glasdegib is resumed after HSCT) are not considered "new anti-cancer drug therapy". Adverse events occurring on the same day as the first dose of study treatment will be considered to have occurred during the on-treatment period. All other assessments which occur on the same day as the first dose of study treatment will be considered baseline assessments (see Section 3.4 for the definition of baseline).

Safety data collected outside the on-treatment period as described above will be listed but not summarized.

3.5.1. Adverse Events

An adverse event is considered treatment emergent (TEAE) relative to a given treatment if:

- the event occurs during the on-treatment period (regardless if it was seen prior to the start of treatment)

CCI [redacted]

Adverse Events of Special Interest (AEoSI)

The current full list of AEoSI and MedDRA high level terms (HLT), standardized MedDRA queries (SMQ), and/or preferred terms (PTs) included in each event are listed in a separate AEoSI cluster document. These events will be defined based on a list of MedDRA Preferred Terms specified in the Safety Review Plan for glasdegib. A final list will be provided to programming prior to database release.

3.5.2. Laboratory Data

Hematology, chemistry, lipid and coagulation tests results will be programmatically graded according to the NCI CTCAE version 4.03 for relevant parameters. Parameters which cannot be graded will be summarized relative to the normal range (ie, normal range high or normal range low). Additional details are provided in Section 6.6.3.

4. ANALYSIS SETS

Data for all patients will be assessed to determine if patients meet the criteria for inclusion in each analysis population prior to unblinding and releasing the database and classifications will be documented per standard operating procedures. The intensive and non-intensive chemotherapy patients are always analyzed separately.

4.1. Full Analysis Set

The Full Analysis (FA) set will include all randomized subjects. Subjects will be classified according to the treatment assigned at randomization.

Randomized but not treated patients will be reported under their randomized treatment group for the full analysis set. The FA set will be the primary population for evaluating all efficacy and PRO endpoints CCI and patient characteristics.

4.2. Per Protocol Analysis Set

Not applicable.

4.3. Safety Analysis Set

The Safety Analysis (SA) set will include all subjects who receive at least one dose of study drug.

A randomized but not treated patient will be excluded from this safety analysis set. Subjects will be classified according to the treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case subjects will be classified according to the first study treatment received. The SA set will be the primary population for evaluating treatment administration/compliance and safety, CCI . Efficacy endpoints may be assessed in this population as well.

4.4. Other Analysis Sets

4.4.1. Pharmacokinetics Analysis Set

The PK concentration analysis set is defined as all subjects who are treated and who have at least 1 value of analyte concentration of glasdegib available. The PK parameter analysis set is defined as all subjects who are treated and who have at least 1 of the PK parameters of interest. Subjects will be classified according to the treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case subjects will be classified according to the first study treatment received.

CCI [redacted]

4.4.3. China Sub-Population Analysis Sets

For each efficacy, safety, CCI [redacted] and pharmacokinetics analysis set defined above, the corresponding China subset analysis set is the subset of patients consisting of patients enrolled in mainland China.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and Sample Size Section

The hypotheses and sample size calculation were done independently and separately for the patients in the two studies.

Intensive Chemotherapy Study:

The primary objective of the study is to demonstrate that glasdegib in combination with 7+3 (Arm A) is superior to placebo in combination with 7+3 (Arm B) in prolonging OS.

The study is designed to test H_0: HR ≥1 vs. H_A: HR <1, where HR is the hazard ratio (Arm A/Arm B) of OS.

A total of 400 subjects eligible to receive intensive chemotherapy per investigator assessment will be enrolled into the study. A total of 267 death events to be observed in the study would provide 90% power to detect an improvement in overall survival translated in mOS from 21 months to 31.5 months assuming a composite mOS for the population as follows: young AML (aged ≤60 years) with a mOS of 23.7 months and elderly AML (aged >60 years) with a mOS of 15 months assuming a 70% young and 30% elderly AML split[12,13] and with an HR=0.67 using a 1-sided log-rank test at a significance level of 0.025 and a 3-look group-sequential design.

The sample size of 400 subjects is determined based on the assumption of a HR of 0.67 under the alternative hypothesis (under an exponential model, translating in median OS of 31.5 months in the treatment arm and 21 months in the control arm) expecting approximately one third of all enrolled subjects to be censored. It is assumed a ramping enrollment of 10 months to reach a maximum enrollment of 18 patients/month. An accrual duration of 28 months to randomize 400 patients and an overall trial duration of 58 months is expected.

For the China subset, no formal statistical testing will be conducted.

Non-Intensive Chemotherapy Study:

The primary objective of the study is to demonstrate that glasdegib in combination with azacitidine (Arm A) is superior to placebo in combination with azacitidine (Arm B) in prolonging OS.

The study is designed to test H_0: HR ≥1 vs. H_A: HR <1, where HR is the hazard ratio (Arm A/Arm B) of OS.

A total of 320 subjects who are not candidates to receive Intensive Chemotherapy will be enrolled into the study. A total of 220 death events to be observed in the study would provide 90% power to detect an improvement in overall survival translated in mOS from 10.4 months[14] to 16.2 months with an HR=0.64 using a 1-sided log-rank test at a significance level of 0.025 and a 2-look group-sequential design. One interim efficacy and futility analysis of OS would be planned after 60% death events occur in the non-intensive chemotherapy population or upon completion of enrollment of these subjects, whichever is later.

The sample size of 320 subjects is determined based on the assumption of a HR of 0.64 under the alternative hypothesis (under an exponential model, translating in median OS of 16.2 months in the treatment arm and 10.4 months in the control arm) expecting approximately 31% of all enrolled subjects to be censored. It is assumed a ramping enrollment of 10 months to reach a maximum enrollment of 18 patients/month. An accrual duration of 23 months to randomize 320 patients and an overall trial duration of 37 months is expected.

For the China subset, no formal statistical testing will be conducted.

5.1.2. Decision Rules

The interim analyses will be performed based on the FA set for the intensive and non-intensive chemotherapy populations separately and independently of each other as described in the following. The key secondary Fatigue endpoint will be analyzed using a hierarchical testing procedure, provided the primary OS endpoint is statistically significant favoring the experimental arm within each patient population (intensive and/or non-intensive).

An interim analysis will not be performed for the China subset. Full analysis following early stopping at an interim analysis or following final analysis will be carried out for the China subset based on the specifications in other sections.

Non-binding for the futility implies that the futility boundary will be constructed in such a way that it can be overruled if desired by the external data monitoring committee (E-DMC) without inflating the type-1 error rate and without decreasing the power.

Intensive Chemotherapy Study:

Two interim analyses are planned. One efficacy and futility interim analysis is planned when 50% of death events have occurred in the intensive chemotherapy population. However, no efficacy stopping will be made even if the interim analysis result crosses the efficacy boundary. A second interim analysis of efficacy and futility is planned when 70% of death events have occurred or upon completion of enrollment of the intensive chemotherapy population, whichever is later. An O'Brien-Fleming boundary with Lan-DeMets spending function will be used for efficacy and a Rho(3) β-spending function will be used for the non-binding futility boundary. If exactly 50% of the events is accrued at the first interim analysis, the futility boundary will be crossed when 1-sided $p>0.461$. If exactly 70% of the events is accrued at the second interim analysis, the futility boundary will be crossed when 1-sided $p>0.186$ and the efficacy boundary will be crossed when 1-sided $p<0.007$. If the actual number of events at the interim analyses is different, the corresponding spending function will be used to calculate the actual stopping boundaries. A 1-sided $p<0.023$ at the final analysis suggests the null hypothesis is to be rejected.

Table 2. Planned Stopping Boundaries for Overall Survival at Interim Analyses and Final Analysis for the Intensive Chemotherapy Study

Analysis	Number of Events (%)	Cumulative Alpha Spent	Cumulative Beta Spent	Futility Boundary p-value	Efficacy Boundary p-value
First IA	134 (50%)	0.002	0.013	0.461	*
Second IA	187 (70%)	0.007	0.034	0.186	0.007
FA	267 (100%)	0.025	0.100	0.023	0.023

Abbreviations: IA=interim analysis; FA=final analysis
The cumulative alpha spent, cumulative beta spent, and boundaries were calculated using software EAST (version 6.4).
* No stopping for efficacy regardless whether the efficacy boundary is crossed or not.

Non-intensive Chemotherapy Study:

One interim analysis is planned. An interim analysis of efficacy and futility is planned when 60% of death events have occurred or upon completion of enrollment of the non-intensive chemotherapy population, whichever is later. An O'Brien-Fleming boundary with Lan-DeMets spending function will be used for efficacy and a Rho(3) β-spending function will be used for the non-binding futility boundary. If exactly 60% of the events is accrued at the interim analysis, the futility boundary will be crossed when 1-sided $p>0.302$ and the efficacy boundary will be crossed when 1-sided $p<0.004$. If the actual number of events at the interim analyses is different, the corresponding spending function will be used to calculate the actual stopping boundaries. A 1-sided $p<0.024$ at the final analysis suggests the null hypothesis is to be rejected.

Table 3. Planned Stopping Boundaries for Overall Survival at Interim Analyses and Final Analysis for the Non-Intensive Chemotherapy Study

Analysis	Number of Events (%)	Cumulative Alpha Spent	Cumulative Beta Spent	Futility Boundary p-value	Efficacy Boundary p-value
IA	132 (60%)	0.004	0.021	0.302	0.004
FA	220 (100%)	0.025	0.100	0.024	0.024

Abbreviations: IA=interim analysis; FA=final analysis
The cumulative alpha spent, cumulative beta spent, and boundaries were calculated using software EAST (version 6.4).

5.2. General Methods

Unless otherwise specified, baseline data will be summarized by treatment arm and for both treatment arms combined. Disposition, efficacy, exposure (including concomitant therapies), and safety data will be summarized by treatment arm only. The intensive and non-intensive chemotherapy patients will be analyzed separately.

5.2.1. Pooling of Data by Center

Each study is expected to enrol in more than 20 countries and more than 100 sites. In order to provide overall estimates of treatment effects, data will be pooled across centers. The 'center' factor will not be considered in statistical models or for subset analyses due to the high number of participating centers in contrast to the anticipated small number of patients randomized/treated at each center.

No pooling across centers will be done for the China subset analysis.

5.2.2. Definition of Study Day

Start day of study treatment is the day of the first dose of study treatment.

The study day for assessments occurring on or after the first dose of study treatment (eg, adverse event onset, laboratory date) will be calculated as:

Study day = Date of the assessment/event - start date of study treatment + 1.

The study day for assessments occurring prior to the first dose of study treatment (eg, baseline characteristics, medical history) will be negative and calculated as:

Study day = Date of the assessment/event –start date of study treatment.

The study day will be displayed in all relevant data listings.

5.2.3. Definition of Cycle and Cycle Day

Non-intensive Chemotherapy Study

Cycle start and end dates are derived per patient and not per study treatment.

- For Cycle X, the actual cycle start date for each subject is the earliest start date of dosing (dose>0 at that visit) in Cycle X visit CRF exposure page.
- For all but the last cycle.
    - actual cycle end date is calculated as the start date of the next cycle minus one day.
    - actual cycle duration is calculated from Day 1 of a cycle to the day prior to Day 1 of the next cycle, as follows:

        Actual Cycle Duration = cycle end date – cycle start date + 1.
- For the last cycle, actual cycle duration is calculated as follows:
    - Actual Cycle Duration = last date of non-zero study treatment– cycle start date + 1.

The cycle day will be calculated as:

Cycle day = Date of the assessment/event – cycle start date + 1.

Intensive Chemotherapy Study

Study day in association with study period will be displayed for intensive chemotherapy patients, ie, Induction 1, Induction 2 (optional), Cytarabine Consolidation Cycle 1 through 4 (optional), HSCT period (optional), Post-consolidation Cycle 1, 2, 3, etc (optional). Note that a patient who undergoes consolidation therapy can either go on Cytarabine Consolidation and/or HSCT.

Cycle start and end dates are derived per patient and not per study treatment.

For all cycles other than the HSCT period.

- For any cycle, the actual cycle start date for each subject is the earliest start date of dosing (dose >0 at that visit) in that cycle CRF exposure page.

- For all but the last cycle.
    - actual cycle end date is calculated as the start date of the next cycle minus one day.
    - actual cycle duration is calculated from Day 1 of a cycle to the day prior to Day 1 of the next cycle, as follows:

        Actual Cycle Duration = cycle end date – cycle start date + 1.

- For the last cycle, actual cycle duration is calculated as follows:
    - Actual Cycle Duration = last date of non-zero study treatment– cycle start date + 1.

The cycle day will be calculated as:

Cycle day = Date of the assessment/event – cycle start date + 1.

HSCT Period

The HSCT period is defined as time from the start of the HSCT conditioning regimen until the subject re-starts glasdegib/placebo or cytarabine, death, post study follow-up, or withdrawal. Day 1 is the day of the HSCT. Note that in the Protocol this was referred to as Day 0 per common clinical practice. Per Protocol Section 8, there are limited data collection during the HSCT period beyond 28 days after the last investigational treatment administration prior to HSCT.

The cycle day for assessments occurring on or after the day of HSCT will be calculated as:

Cycle day = Date of the assessment/event – date of HSCT + 1.

The cycle day for HSCT conditioning regimen(s) will be negative as they occur prior to HSCT and calculated as:

Cycle day = Date of the assessment/event – date of HSCT.

5.2.4. Date of Last Contact

The date of last contact will be derived for patients not known to have died at the data cutoff date using the latest complete date (ie, imputed dates will not be used in the derivation) among the following:

- All patient assessment dates (eg, blood draws (laboratory, Pharmacokinetics (PK)), vital signs, physical exam, performance status, ECG, Echocardiograms (ECHO)/MUGA scans, disease assessments).
- Start and stop dates of concomitant therapies including non-drug treatments or procedures.

- Completion dates for PRO Questionnaires.
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation including systemic therapy, radiation, and surgeries.
- AE start and end dates.
- Last date known to be alive on the 'Survival Follow-up' CRF.
- Study treatment start and end dates.
- Hospitalization admission and discharge dates.
- Randomization date, and
- Date of discontinuation on disposition CRF pages (do not use if reason for discontinuation is lost to follow-up or death).

Only dates associated with actual examinations of the patient will be used in the derivation. Dates associated with a technical operation unrelated to patient status such as the date a blood sample was processed or dates data were entered into the CRF will not be used. Assessment dates after the data cutoff date will not be applied to derive the last contact date.

5.2.5. Disease Assessment Date

Date of assessment from the hematologic response CRF will be used as the date of response. Date of hematologic relapse will be derived as the earliest date from the peripheral blood blasts, bone marrow blasts, or EMD assessment where the criteria for relapse is met. For hematologic relapse occurring during long-term follow-up, the date of assessment from the hematologic response CRF will be used.

5.2.6. Unscheduled Assessments

Unless otherwise specified, unscheduled assessments will not be displayed in summary tables by nominal visit/timepoint. Unscheduled assessments will be used when deriving baseline and worst case on-treatment for safety and PRO analyses (except where noted for baseline ECGs). Additionally, unscheduled disease assessments will be used for efficacy analyses (eg, defining date of relapse/censoring, CRi or better, date of last contact).

5.2.7. Standard Derivations and Reporting Conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

Percentages will be reported to one decimal place. The rounding will be performed to closest integer/first decimal using the common mid-point between the two consecutive values, eg, 5.1 to 5.4 will be rounded to an integer of 5, and 5.5 to 5.9 will be rounded to an integer of 6.

5.2.8. Analyses for Continuous Data

Continuous variables will be summarized using descriptive statistics ie, number of non-missing values and number of missing values [ie, n (missing)], mean, median, standard deviation (SD), minimum, maximum and first and third quartile (Q1 and Q3). CCI [redacted] PK summaries will also include coefficient of variation percent (%CV).

In case the analysis refers only to certain visits, estimates will be based on the number of patients with an assessment at that visit, unless otherwise specified.

5.2.9. Analyses for Categorical Data

Qualitative variables will be summarized by frequency counts and percentages. Unless otherwise specified, the calculation of proportions will include the missing category. Therefore counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of patients with an assessment at that visit, unless otherwise specified.

5.2.10. Analyses for Time to Event Data

The stratified log-rank test will be used for comparing treatments. Hazard ratios and the associated 95% two-sided confidence intervals are estimated by Cox proportional hazards (PH) model. Repeated confidence interval will also be provided, to account for alpha spending, when appropriate.

Time to event endpoints will be summarized using the Kaplan-Meier method and estimated survival curves will be displayed graphically when appropriate. Graphs will describe the number of patients at risk over time. The median, first and third quartiles, and probabilities of an event at particular points in time will be estimated by the Kaplan-Meier method. Confidence intervals for medians and quartiles are based on the Brookmeyer-Crowley[2] method. Confidence intervals for the estimated probability of an event at a particular timepoint will be generated using the log(-log) method with back transformation to a confidence interval on the untransformed scale. Summaries of the number and percentages of patients with an event and reason for censoring will also be provided on summary tables and/or figures.

If a cumulative incidence analysis is performed, the stratified Gray's test[17] will be used for comparing treatments. Hazard ratios and the associated 95% two-sided confidence intervals are estimated by a proportional subdistribution hazards model. The 25^{th}, 50^{th}, and 75^{th} percentiles and probabilities of an event at particular points in time will be estimated from the cumulative incidence curve. Confidence intervals for the estimated probability of an event at a particular timepoint will be generated using the delta method with a log(-log) transformation for the standard error.

5.3. Methods to Manage Missing Data

Unless otherwise specified, all data will be evaluated as observed, and no imputation method for missing values will be used.

Any imputations will occur at the analysis dataset level. Additionally, in all patient data listings imputed values will be presented and flagged as imputed.

Missing statistics, eg, when they cannot be calculated, should be presented as 'ND' for not done, 'NR' for not reached or 'NA' for not applicable. For example, if N=1, the measure of variability cannot be computed and should be presented as 'ND' or 'NA'.

CDISC Safety Rulebook will be followed when handling the missing/partial dates, treatment emergent AE algorithm, missing AE grades, etc.

5.3.1. Missing Dates

For purposes of data listings, dates will reflect only the information provided by the investigator on the CRF.

If start dates for adverse events or concomitant medications are completely missing a worst case approach will be taken whereby the events will be considered treatment emergent and the medications will be considered concomitant. If only partial information are available (eg, only a month and year or only a year) and the partial information provide sufficient information to indicate the dates are prior to the start of study treatment (eg, month/year less than month/year of first dose) then these will be considered to have started prior to treatment; otherwise a similar worst case approach will apply and these will be considered to have started after treatment.

Date of Last Dose of Study Treatment

No imputation will be done for first dose date. Date of last dose of study treatment, if unknown or partially unknown, will be imputed as follows:

- If the last date of study treatment is completely missing and there is no Disposition CRF page for the treatment phase and no death date, the patient should be considered to be ongoing and use the data cutoff date for the analysis as the last dosing date; or

- If the last date of study treatment is completely or partially missing and there is EITHER a Disposition CRF page for the treatment phase OR a death date available (on or prior to the data cutoff date), then impute this date as the last dose date:

  = 31DECYYYY, if only Year is available and Year < Year of min (Date of Completion/Discontinuation from the Disposition CRF page for the treatment phase, death date),

  = Last day of the month, if both Year and Month are available and Year = Year of min (Date of Completion/Discontinuation from the Disposition CRF page for the treatment phase, death date) and Month < the month of min (EOT date, death date), or

  = min (Date of Completion/Discontinuation from the Disposition CRF page for the treatment phase, death date), for all other cases.

Missing or Partial Death Dates

Missing or partial death dates will be imputed based on the last contact date:

- If the entire date is missing it will be imputed as the day after the date of last contact (see derivation of date of last contact in Section 5.2.4); or
- If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) day after the date of last contact and the following:
  - Missing day: 1st day of the month and year of death, or
  - Missing day and month: January 1st of the year of death.

Date of Start of New Anti-cancer Therapy

Incomplete dates for new anti-cancer therapy will be imputed as follows and will be used to determine censoring dates for efficacy analyses. Note that HSCT before the end of the treatment phase or its conditioning regimen(s) are not considered new anti-cancer therapies.

- The end date of new anti-cancer therapy will be included in the imputation for start date of new anti-cancer therapy. If the end data of new anti-cancer therapy is:
  - completely missing then it will be ignored in the imputations below,
  - partially missing with only year available then the imputations below will consider 31DECYYYY as the end date of the new anti-cancer therapy, or
  - partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anti-cancer therapy.
- For patients who have not discontinued study treatment at the time of the data cutoff date, last dose of study treatment is set to the data cutoff date in the imputations below.
- If the start date of new anti-cancer therapy is completely or partially missing then the imputed start date of new anti-cancer therapy is:

  = 31DECYYYY, if only Year is available and Year < Year of min [max (PD/relapse date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

  = Last day of the month, if both Year and Month are available and

  Year = Year of min [max (PD/relapse date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

Month < Month of min [max (PD/relapse date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

= min [max (PD/relapse date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], for all other cases.

AE Onset Date:

The following imputation rules apply if the event is unique for a patient or it is the first of a series of similar events; otherwise, the AE Onset Date will not be imputed:

- If the AE Collection Date is not missing, is less than the Date of First Exposure to Treatment, and is less than the AE Stop Date, then AE Onset Date is set to the Date of AE Collection.
- If the Previous Visit Date is greater than the Date of First Exposure to Treatment and less than the AE Stop Date, the AE Start Date is set to the previous visit date.
- If the Date of First Exposure to Treatment is greater than the previous visit date and less than the AE Stop Date, the AE Onset Date is set to the Date of First Exposure to Treatment.
- Otherwise AE Onset date is set to the AE Stop date.

AE Stop Date:

Ongoing events will have the AE Stop Date set to one of the following values:

- Date of Death, if the patient died and a date of death exists.
- Maximum of (Patient Withdraw date, AE Onset Date, AE Collection Date) if the patient withdrew from the study and a date of withdraw exists.
- Maximum of (AE Onset Date, Subject Summary Collection Date, AE Collection Date) if the Disposition CRF page for the long-term follow-up phase exists but a date of completion/discontinuation does not exist.
- Maximum of (Last Treatment Date, AE Onset Date) if no Disposition CRF page for the long-term follow-up phase exists.

Imputation will only occur if event is unique for the patient, or it is the last of a series of similar events; otherwise the Stop Date will not be imputed. Adverse Events are deemed similar if they have the same verbatim term.

Resolved events will have the AE Stop Date set to the maximum of the AE collection date and the AE Onset date.

Other Missing or Partial Dates

Imputation methods for other partial dates are as follows:

- If the day of the month is missing for a start date used in a calculation, the first day of the month will be used to replace the missing date.
- If both the day and month are missing for a start date, the first day of the year is used.
- For stop dates, the last day of the month, or last day of the year is used if the day or day and month are missing, respectively.
- If the date is completely missing, no imputation will be performed.

5.3.2. Missing Toxicity Grade of Adverse Events

During Study Treatment: If no toxicity grade is available or the grade is reported as unknown for an adverse event during the study treatment, then the event will be considered treatment emergent.

In summaries which present maximum toxicity grade, the maximum of non-missing grades will be displayed. Missing grade will only be displayed for cases where a patient reported only one event and the grade is missing.

5.3.3. Missing Pharmacokinetic (PK) Data

Concentrations below the limit of quantification

For all calculations, figures, and estimation of individual pharmacokinetic parameters, all concentrations assayed as below the level of quantification (BLQ) will be set to zero. In log-linear plots these values will not be represented. The BLQ values will be excluded from calculations of geometric means and their confidence intervals. A statement similar to 'All values reported as BLQ have been replaced with zero' should be included as a footnote to the appropriate tables and figures. In listings BLQ values will be reported as below limit of quantification ("<LLOQ"), where LLOQ will be replaced with the corresponding value from the analytical assay used.

Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of median profiles, concentrations will be set to missing if one of the following cases is true:

- A concentration has been reported as ND (ie, not done) or NS (ie, no sample);
- A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the clinical pharmacologist.

Summary statistics will not be presented at a particular timepoint if more than 50% of the data are missing. For analysis of pharmacokinetic concentrations, no values will be imputed for missing data.

5.3.4. Missing ECG Data

For ECG analyses, no values will be imputed for missing data (except HR values that can be derived from RR values, if present. If both RR and HR values are missing, QTcB and QTcF will not be determined). In case of missing HR value, HR (bpm) will be derived as (60/RR [sec]), if RR is collected.

If one or two of the triplicate measurements for an ECG parameter are missed, the average of the remaining two measurements or the single measurement can be used in the analyses. If all triplicate measurements are missing at a timepoint for an ECG parameter, no values will be imputed for this timepoint. If the triplicate needs to be repeated because of an artifact, then the repeated triplicate will be reported on an unscheduled CRF page. Based on a review of the data these unscheduled assessments may be used in place of the assessments at the nominal time. Data review and consultation with the study team is required to flag these cases.

CCI

[redacted]

6. ANALYSES AND SUMMARIES

Analyses for the intensive and non-intensive chemotherapy studies will be conducted separately and independently. Unless otherwise stated, stratified analysis will be stratified based on ELN genetic risk, age ≤60 vs >60 years, and region (rest of world [ROW] vs China) for the intensive patients, and ELN genetic risk and age <75 vs ≥75 years for the non-intensive patients. The rate of stem cell transplantation is expected to be much lower for sites in China versus the ROW thus region is being added as a stratification factor for the intensive study analyses. The rate of stem cell transplantation is expected to be low in all regions in the non-intensive study.

6.1. Primary Endpoint - Overall Survival

The primary efficacy analysis will compare OS between the experimental arm and the control arm, and will be performed using a one-sided stratified log-rank test. The primary population for the analysis will be the FA set. Strata will be based on those specified in the randomization system including region for the intensive study.

OS is defined as the time from randomization to the date of death due to any cause. Patients last known to be alive will be censored at date of last contact (see Section 5.2.4). OS will be summarized in months:

OS (months) = [date of death or censoring–randomization+1]/30.4375.

The treatment effect will be estimated using a Cox's PH model stratified by the randomization strata and region for the intensive study to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie, for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=control arm, 1= experimental arm) and β is the unknown regression parameter. Ties will be handled using the Efron method as an approximation to the exact method (Ties=EFRON option in SAS PROC PHREG).

OS time associated with each treatment arm will be summarized using the Kaplan-Meier method (product-limit estimates) and displayed graphically where appropriate. CIs for the 25^{th}, 50^{th}, and 75^{th} percentiles will be reported. Repeated confidence interval will also be provided, to account for alpha spending. The Cox PH model will be fitted to compute the treatment HRs and the corresponding 95% CIs. In addition, the median OS and its two-sided 95% CI using the Brookmeyer and Crowley method will be provided for each stratum within each treatment arm separately. The HR and its two-sided 95% CI will be provided for each stratum as well.

The OS rate at 12, 24 and 36 months will be estimated with corresponding two-sided 95% CIs for the intensive chemotherapy patients. The OS rate at 6, 12, and 18 months will be estimated with corresponding two-sided 95% CI for the non-intensive chemotherapy patients. The CIs for the median will be calculated according to Brookmeyer and Crowley[2] and the CIs for the survival function estimates at the timepoints defined above will be derived using the log(-log) method according to Kalbfleisch and Prentice[3] (conftype=loglog default option in SAS Proc LIFETEST). The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with an event (death) and censoring reasons will be presented by treatment arm. Censoring reasons are as follows:

- Alive;
- Withdrawal of consent;
- Lost to follow-up.

The OS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Should over-stratification prevent model convergence or there exists large imbalance in the distribution (eg, 90% or more subjects from one strata level), the intermediate and adverse ELN risk groups may be pooled in the analysis.

Time of Follow-Up for OS

A Kaplan-Meier summary table for OS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the OS censoring and event indicators.[15]

CCI

[redacted]

CCI

[redacted]

6.2. Secondary Endpoints

6.2.1. Fatigue

The "Fatigue" single-item from the MDASI-AML/MDS questionnaire is the key secondary endpoint. Similar to analysis of the primary endpoint, the analysis of Fatigue will be conducted separately and independently across the two patient populations (intensive or non-intensive patients). There are two steps of the analysis of Fatigue. If the primary endpoint of OS is met in one population, the hypothesis testing described in Step 2 of this section will be carried out in that population. The overall Type I error is preserved by following this pre-specified hierarchical testing sequence (ie, OS followed by Fatigue).

Step 1: A repeated measures model will be used to determine Clinically Important Responder (CIR) within each patient population. We use the change in Fatigue from baseline as the outcome and SGIC-S (Subject Global Impression of Change using PGIS) score (defined below) as the anchor. Note that Fatigue and PGIS are collected following the same schedule and treatment assignment is not a covariate in the model. All available post-baseline data up to Week X will be used (X=8 for intensive patients, X=12 for non-intensive patients). A window of +/-3 days will be applied. If more than one assessment are available in a week, the average value will be used.

A SGIC-S score with three possible values (-1, 0, and 1, treated as continuous variable) will be derived for each subject at each week post-baseline, from Week 1 through Week X, as follows:

a. SGIC-S = -1, if post-baseline PGIS minus PGIS at baseline is positive (this means ***worse***);

b. SGIC-S = 0, if post-baseline PGIS minus PGIS at baseline is zero (ie, no change, this means ***the same***);

c. SGIC-S = 1, if post-baseline PGIS minus PGIS at baseline is negative (this means ***better***).

The predicted change in Fatigue score corresponding to one unit difference on SGIC-S is the threshold for "responder" vs "non-responder". All patients with change in Fatigue at Week X (X=8 for intensive patients, X=12 for non-intensive patients) at or below the threshold (note that lower values suggest lessening of Fatigue) will be defined as responders.

As a supportive analysis, a SGIC-C (Subject Global Impression of Change using PGIC) score with three possible values (-1, 0, and 1, treated as continuous variable), derived below, will be used in place of SGIC-S in the repeated measures model. Note that PGIC is collected on the same schedule as Fatigue other than there is no assessment of PGIC at baseline.

a. SGIC-C = -1, if PGIC is 5 or more (this means ***worse***);

b. SGIC-C = 0, if PGIC is 4 (ie, no change, this means ***the same***);

c. SGIC-C = 1, if PGIC is 3 or less (this means ***better***).

Additional details of the CIR can be found in the Appendix, where potential sensitivity analyses treating SGIC-S and SGIC-C as categorical variables are also described.

Step 2: A formal testing of proportion of responders at Week X (using a 1-sided Cochran Mantel Haenszel [CMH] test, stratified by ELN risk, age [age $\leq$60 vs >60 years for the intensive patients, and age <75 vs $\geq$75 years for the non-intensive patients], and region [ROW vs China for intensive patients only) will only be conducted in one population if the primary OS endpoint is met in that population (intensive or non-intensive patients). The risk ratio and 2-sided 95% CI will be summarized.

The proportion of responders by treatment arm will be estimated with a 2-sided 95% CI (using normal approximation). The proportion and 2-sided 95% CI (using exact method) of responders for each stratum by treatment arm will also be provided.

In addition to the above responder analyses, analysis of Fatigue will also be supplemented with cumulative distribution function (CDF) and probability density function (PDF) of change in Fatigue score from baseline to Week X. The change in Fatigue at Week X will be plotted on the X-axis while its CDF or PDF will be plotted on the Y-axis.

If the Week X PRO assessment is missing, then the next available completed PRO assessment will be used.

6.2.2. Responses

The proportion of patients achieving $CR_{MRD\text{-}negative}$, CR (including $CR_{MRD\text{-}negative}$), CRi, CRh (non-intensive only), MLFS, and PR as their best overall response will be estimated with two-sided 95% CI (using normal approximation). The proportion and two-sided 95% CI (using exact method) of patients achieving each response category for each stratum will also be provided. The proportion of patients with $CR_{MRD\text{-}negative}$, CR (including $CR_{MRD\text{-}negative}$), CRi or better and CRh or better (non-intensive only) respectively will be compared between the 2 treatment arms using a 1-sided CMH stratified test and an unstratified chi-square test. The risk ratio and difference in proportions and their 95% CIs will be provided overall and within each stratum as well. The proportion of patients experiencing stable disease, indeterminate response (eg, those that met the criteria for stable disease but for <3 months in the non-intensive study or those where response could never be assessed due to missing information such as peripheral or bone marrow blasts) and disease progression as their best response or ever experiencing disease progression will also be summarized. For those patients not assessed for response, the reason for treatment discontinuation will be summarized. Patients with indeterminate as their best response will be categorized into one of the following groups: no post-baseline bone marrow blasts available, post-baseline bone marrow blasts available but missing peripheral blood data, stable disease for <3 months in the non-intensive study, or other. Patients without a response CRF will be classified as Not Evaluated and the overall reason for discontinuation from treatment will be summarized.

6.2.3. Duration of Response

Duration of Response (DoRi or DoRh [non-intensive only]) is defined, for patients from the FA set with a best overall response of CRi or better or CRh or better, as the time from first documentation of response (CRi or better or CRh or better) to the date of first documentation of objective disease relapse or death due to any cause. The censoring rules for DoRi/DoRh are discussed Section 3.2.1.

DoRi (months) = [date of event or censoring– first date of CRi or better +1]/30.4375.

DoRh (months) = [date of event or censoring– first date of CRh or better +1]/30.4375.

Depending on the number of subjects who achieve a best overall response of CRi or better or CRh or better and subsequently have an event in each arm, Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median DoRi or DoRh (non-intensive only) time with two-sided 95% CIs. Inferential analyses (eg, a hazard ratio and p-value) will not be provided for duration of response.

Duration of response for the subset of patients with a best overall response of CR or better will also be analyzed.

An alternative definition of DoRi defined for all FA patients, may be explored for the intensive chemotherapy patients, where patients who have never achieved CRi or better will be considered having an event on Day 1.[16]

6.2.4. Time to Response (Non-intensive Only)

Time to Response (TTRi or TTRh) is defined, for patients achieved CRi or better or CRh or better, as the time from the date of randomization to the first documentation of response (CRi or better or CRh or better).

The minimum, median, and maximum TTRi/TTRh for each treatment arm will be provided.

6.2.5. Event-free Survival

A 1-sided stratified log-rank test (stratified by ELN genetic risk, age [age ≤60 vs >60 years, and region [ROW vs China] for the intensive patients, and genetic risk and age <75 vs ≥75 years for the non-intensive patients]) will be used. Analyses will be performed using various EFS definitions as specified in Section 3.2.1. Should over-stratification prevent model convergence or there exists a large imbalance in the distribution (eg, 90% or more subjects from one strata level), the intermediate and adverse ELN risk groups may be pooled in the analysis. EFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and displayed graphically where appropriate. CIs for the 25th, 50th, and 75th percentiles of EFS probabilities will be reported (using Brookmeyer and Crowley method). In addition, the median EFS and its two-sided 95% CI using the same method will be provided for each stratum within each treatment arm separately. The Cox PH model will be fitted to compute the treatment HR and the corresponding 95% CI. The HR and its two-sided 95% CI will be provided for each stratum as well. The proportion of patients experiencing each type of EFS event (eg, TF, hematologic relapse, death) will also be summarized.

6.2.6. Patient-Reported Outcomes

All PRO analyses will be based on the FA[redacted].

6.2.6.1. MDASI-AML/MDS

A MDASI-AML/MDS completion table will be constructed showing for each treatment group at each time point the number and percentage of subjects who completed and who had not completed the questionnaires, and also the reasons for non-completion.

Analysis of MDASI-AML/MDS will be based on the core cancer symptom score, the AML/MDS-specific symptoms score, the interference areas score, and also the 5 individual items of fatigue, disturbed sleep, dry mouth, muscle weakness, and lack of appetite. Note that Fatigue also appears in Section 6.2.1 (as a key secondary endpoint) using a responder analysis.

A display of descriptive statistics including means, medians, standard deviations, and 95% confidence intervals at each assessment point will be provided. This will be done based on the observed values as well as on change from baseline values. Statistical comparison of the 2 treatment groups will be based on a longitudinal repeated measures mixed effects model using baseline as a covariate.

A graphical representation of observed mean and standard error over time and change from baseline over time by treatment arms will also be provided for the core cancer symptom score, the AML/MDS-specific symptom score, the interference area score, and also the 5 individual items of fatigue, disturbed sleep, dry mouth, muscle weakness, and lack of appetite.

6.2.6.2. EQ-5D-5L

An EQ-5D completion table will be constructed showing for each treatment group at each time point the number and percentage of subjects who completed and who had not completed the questionnaires, and also the reasons for non-completion.

Analysis of EQ-5D-5L and EQ-VAS will follow similar methodology as for the MDASI endpoints described above. It will include a display of descriptive statistics including means, medians, standard deviations, and 95% confidence intervals at each assessment point. This will be done based on the observed values as well as on change from baseline values. Statistical comparison of the 2 treatment groups on EQ-5D and EQ-VAS will be carried out using a longitudinal repeated measures mixed effects model with baseline as a covariate.

In addition to the above analyses, a health status profile table showing the number of patients responding at each level on each dimension at each timepoint will also be constructed.

Finally as with the MDASI variables, a graphical representation of observed mean and standard error over time and change from baseline over time by treatment arms for EQ-5D and EQ-VAS will also be provided.

6.2.6.3. PGIS

PGIS will be used as an anchor for defining “responder” versus “non-responder” for MDASI-AML/MDS individual item Fatigue. Analysis method for anchoring is given in Section 6.2.1 under analysis for Fatigue. A frequency distribution of PGIS categories at each time point by treatment arm will also be provided.

6.2.6.4. PGIC

As with PGIS, a frequency distribution of PGIC categories by treatment arm will be provided and PGIC will also serve as a secondary anchor to define a responder definition for MDASI-AML/MDS individual item Fatigue and as a sensitivity analysis for defining a “clinically important responder” on this endpoint.

CCI

[redacted]

6.3. Other Endpoints

CCI

[redacted]

CCI

6.3.3. Pharmacokinetic Analysis

PK Parameters:

For intensive and non-intensive studies, the plasma trough concentration (C_{trough}) will be reported. Descriptive statistics will be provided for these PK parameters in tabular form (n, mean, standard deviation (Stdev), CV, median, minimum, maximum, geometric mean and its associated CV) by cycle (or Induction 1, 2) and day.

PK Concentrations:

For drug concentrations, individual values and descriptive statistics (n, mean, Stdev, CV, median, minimum, maximum, geometric mean and its associated CV) will be presented by cycle (or Induction 1, 2), day of assessment, and nominal time in tabular form.

6.3.4. Population Pharmacokinetic Analysis or PK/PDx Modeling

PK and pharmacodynamics (PDx) data from this study may be analyzed using modeling approaches and may also be pooled with data from other studies to investigate any causal relationship between study treatment exposure and biomarkers or significant safety endpoints. The results of these analyses, if performed, may be reported separately.

CCI [redacted]

CCI

[redacted]

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

The following analyses will be based on the FA set overall and separately by treatment arm.

Descriptive baseline summaries by race will not be performed for the China subset.

- **Demographic and Physical Characteristics**

  The following demographic and baseline characteristics will be summarized by number and percentage:

  - Gender (male, female).
  - Age (18-<45; 45- <65; 65-<75; ≥75), additionally ≤60 years vs >60 years for intensive patients and <75 years vs ≥75 years for non-intensive patients.
  - Race (White, Black, Asian, Other), Racial designation for Asian (Japanese, Korean, Chinese, other), ethnicity (Hispanic/Latino, non-Hispanic/Latino).
  - Eastern Cooperative Oncology Group (ECOG) Performance status.

  Age (continuous), height (cm), weight (kg), will be summarized with descriptive statistics (mean, median, standard deviation, minimum, and maximum).

- **Disease Characteristics**

  The following baseline disease characteristics will be summarized by number and percentage:

  - ELN risk category at randomization.
  - Reasons for choosing non-intensive study.
  - Bone marrow blasts (<30, ≥30%).
  - White blood cell (WBC) count (<10, ≥10x10^9/L).
  - Platelet count (<50, ≥50x10^9/L).
  - Hemoglobin (<100, ≥100 g/L).
  - Genetic Abnormality (eg, NPM1 MUTATION WITHOUT FLT3-ITD, etc).
  - Hematological Disease History (de novo, secondary).

  Time from initial diagnosis to screening, peripheral blood blasts and bone marrow blasts (%), WBC (10^9/L), platelet count (10^9/L), and hemoglobin (g/L) will be summarized with descriptive statistics (mean, median, standard deviation, minimum, and maximum).

- **Stratification Factors**

  The number of patients per strata will be summarized based on strata reported in the randomization system as well as on the CRF including a summary of any discordance. The following categories will be considered:

  Intensive Chemotherapy Study

  - age ≤60 years, favorable ELN risk category, and ROW.
  - age ≤60 years, intermediate ELN risk category, and ROW.
  - age ≤60 years, adverse ELN risk category, and ROW.
  - age >60 years, favorable ELN risk category, and ROW.
  - age >60 years, intermediate ELN risk category, and ROW.
  - age >60 years, adverse ELN risk category, and ROW.
  - age ≤60 years, favorable ELN risk category, and Chinese site.

- age ≤60 years, intermediate ELN risk category, and Chinese site.
- age ≤60 years, adverse ELN risk category, and Chinese site.
- age >60 years, favorable ELN risk category, and Chinese site.
- age >60 years, intermediate ELN risk category, and Chinese site.
- age >60 years, adverse ELN risk category, and Chinese site.

Non-intensive Chemotherapy Study

- age <75 years and favorable ELN risk category.
- age <75 years and intermediate ELN risk category.
- age <75 years and adverse ELN risk category.
- age ≥75 years and favorable ELN risk category.
- age ≥75 years and intermediate ELN risk category.
- age ≥75 years and adverse ELN risk category.

- **Medical History**
  - Medical history will be coded using the most current version of MedDRA and summarized by MedDRA's System Organ Class (SOC) and PT. Each patient will be counted only once within each PT or SOC. Summaries will be ordered by primary SOC and PT in descending order of frequency by the experimental treatment arm. Separate summaries will be provided for past and present conditions.

6.5.2. Study Conduct and Patient Disposition

The following analyses will be based on the FA set overall and separately by treatment arm.

6.5.2.1. Patient Disposition

A summary of the number of patients screened and enrolled overall and by country and site will be provided.

Discontinuation of treatment overall and for each study drug, discontinuation of post-treatment follow-up, and discontinuation of study will be summarized separately and by reason for discontinuation. Discontinuations from study treatment and post-treatment follow-up will be summarized using the safety analysis set and discontinuations from the study will be summarized using the FA set. Once a patient discontinues all study drugs they will be considered discontinued overall from study treatment. The reason and date corresponding to the last drug discontinued should be used for the overall summary. If

multiple drugs were discontinued on the same date with different reasons then the reason for discontinuing blinded therapy will be used.

Patients completing blinded therapy will be identified as either those who were confirmed MRD negative and were treated for less than 2 years or those who were treated for 2 years.

Discontinuations from study treatment due to adverse events will be identified as either related or not related to study treatment. If causality is missing the event will be considered related to treatment. If multiple events lead to study treatment discontinuation and at least one was considered related, discontinuation will be reported as related to study treatment.

An additional summary to meet European Union Disclosure requirements will categorize discontinuations due to adverse events based on the following categories:

- Adverse Event, not serious.
- Adverse event, serious non-fatal.
- Adverse event, serious fatal.

6.5.2.2. Protocol Deviations

Protocol deviations will be compiled prior to database closure and will be summarized by category (n(%)) for the FA set by treatment arm. Categories will be assigned by the study Clinician.

6.5.3. Study Treatment Exposure

The following analyses will be based on the SA set.

6.5.3.1. Exposure to Glasdegib

The summary of treatment exposure for glasdegib will include the following information:

- Treatment duration (weeks);
- Cumulative dose (mg);
- Dose intensity (mg/day);
- Relative dose intensity (%).

The duration of glasdegib (in weeks) is defined as:

$$\text{Treatment duration (weeks)} = (\text{last dose date} - \text{first dose date} + 1)/7.$$

The cumulative dose (mg) of glasdegib is the sum of the actual dose levels that the patient received (ie, total dose administered (mg)).

The dose intensity (DI) and the relative dose intensity (RDI) of glasdegib will be calculated for each patient during the study. The DI (mg/day) of glasdegib during the study is defined as:

> DI (mg/day) = [cumulative dose (mg)]/[treatment duration (days) including any dose interruption period prior to permanent treatment discontinuation and excluding the HSCT period, if applicable].

The RDI of glasdegib is defined as the ratio of the DI and planned dose intensity and expressed in %

> RDI (%) = 100× [DI (mg/day]/[100 (mg/day)].

6.5.3.2. Exposure to Cytarabine (Induction, Intensive Study Only)

The summary of treatment exposure for cytarabine in the induction cycles will include the following information:

- Treatment duration (weeks);
- Total number of infusions received;
- Cumulative dose (mg/m^2);
- Dose intensity (mg/m^2/day);
- Relative dose intensity (%).

Cytarabine is administered by intravenous (IV) infusion daily for 7 days, in 28 day cycles (5 days allowed in re-induction cycle).

The dose level for cytarabine captured on the CRF is calculated as actual dose administered/body surface area (BSA) (mg/m^2). The last available weight of the patient on or prior to the day of dosing will be used.

The duration of cytarabine (in weeks) during the study for a patient is defined as:

> Treatment duration (weeks) = (last dose date – first dose date + 1)/7.

The cumulative dose (mg/m^2) of cytarabine per cycle is the sum of the actual dose levels that the patient received within that cycle (ie, total dose administered (mg)/BSA (m^2)). Cumulative dose (mg/m^2) of cytarabine is the sum of the actual dose levels that the patient received throughout induction.

Since the planned cytarabine dose may be 5 (Option 1: ‘5+2’) or 7 (Option 2: ‘7+3’) days in the re-induction cycle, the dose intensity (DI) and the relative dose intensity (RDI) of cytarabine will be calculated for each patient by cycle and across all cycles based on the actual treatment received. The DI (mg/m^2/day) of cytarabine is defined as:

> DI (mg/m^2/day) = [cumulative dose (mg/m^2)]/[number of days with a non-zero dose].

The RDI of cytarabine is defined as the actual DI divided by the planned dose per day and expressed in %:

RDI (%) = 100× [DI (mg/m^2/day)]/[100× (mg/m^2/day)].

6.5.3.3. Exposure to Cytarabine (Consolidation, Intensive Study Only)

The summary of treatment exposure for cytarabine in the consolidation cycles will include the following information:

- Total number of infusions received;
- Cumulative dose (g/m^2);
- Dose intensity (g/m^2/day).

Cytarabine is administered by intravenous (IV) infusion twice daily for 3 days or follow local prescribing information, in 28 day cycles.

The dose level for cytarabine captured on the CRF is calculated as actual dose administered/BSA (g/m^2). The last available weight of the patient on or prior to the day of dosing will be used.

The cumulative dose (g/m^2) of cytarabine per cycle is the sum of the actual dose levels that the patient received within that cycle (ie, total dose administered (g)/BSA (m^2)). Cumulative dose (g/m^2) of cytarabine is the sum of the actual dose levels that the patient received throughout consolidation.

Since the planned cytarabine dose in consolidation cycles can be 1 or 3 g/m^2 BID according to patient age or different following local clinical practice (eg, Japan will use 2 g/m^2 BID for 5 days in a 28 days cycle), the DI of cytarabine will be calculated for each patient by cycle and across all cycles based on the actual treatment received. The DI (g/m^2) of cytarabine is defined as:

DI (g/m^2/day) = [cumulative dose (g/m^2)]/[number of days with a non-zero dose].

The RDI of cytarabine will not be calculated for consolidation since the intended dose is not captured on the CRF.

6.5.3.4. Exposure to Daunorubicin (Induction Only, Intensive Study Only)

The summary of treatment exposure for daunorubicin will include the following information:

- Treatment duration (weeks);
- Total number of infusions received;
- Cumulative dose (mg/m^2);
- Dose intensity (mg/m^2/day);

- Relative dose intensity (%).

Daunorubicin is administered by intravenous (IV) infusion daily for 3 days, in 28 day cycles (2 days allowed in re-induction cycle).

The dose level for daunorubicin captured on the CRF is calculated as actual dose administered/BSA (mg/m^2). The last available weight of the patient on or prior to the day of dosing will be used.

The duration of daunorubicin (in weeks) during the study for a patient is defined as:

Treatment duration (weeks) = (last dose date – first dose date + 1)/7.

The cumulative dose (mg/m^2) of daunorubicin per cycle is the sum of the actual dose levels that the patient received within that cycle (ie, total dose administered (mg)/BSA (m^2)). Cumulative dose (mg/m^2) of daunorubicin is the sum of the actual dose levels that the patient received throughout induction.

Since the planned daunorubicin dose is 2 (Option 1: '5+2') or 3 (Option 2: '7+3') days in the re-induction cycle, the DI and RDI of daunorubicin will be calculated for each patient by cycle and across all cycles based on the actual treatment received. The DI (mg/m^2/day) of daunorubicin is defined as:

DI (mg/m^2/day) = [cumulative dose (mg/m^2)]/[number of days with a non-zero dose].

The RDI of daunorubicin is defined as the actual DI divided by the planned dose per day and expressed in %:

RDI (%) = 100× [DI (mg/ m^2/day]/[60× (mg/m^2/day)].

6.5.3.5. Exposure to Azacitidine (Non-intensive Study Only)

The summary of treatment exposure for azacitidine will include the following information:

- Treatment duration (weeks);
- Total number of injections/infusions received;
- Cumulative dose (mg/m^2);
- Dose intensity (mg/m^2/day);
- Relative dose intensity (%).

Azacitidine is administered by subcutaneous injection (SC) or intravenous (IV) infusion 75 mg/m^2 daily for 7 days, in 28 day cycles.

The dose level for azacitidine captured in the CRF is calculated as actual dose administered/BSA (mg/m^2). The last available weight of the patient on or prior to the day of dosing will be used.

The duration of azacitidine (in weeks) during the study for a patient is defined as:

Treatment duration (weeks) = (last dose date – first dose date + 1)/7.

The cumulative dose (mg/m^2) of azacitidine per cycle is the sum of the actual dose levels that the patient received within that cycle (ie, total dose administered (mg)/BSA (m^2)). Cumulative dose (mg/m^2) of azacitidine is the sum of the actual dose levels that the patient received throughout the study.

The DI) and RDI of azacitidine will be calculated for each patient by cycle and across all cycles. The DI per day (mg/m^2/day) of azacitidine is defined as:

DI (mg/m^2/day) = [cumulative dose (mg/m^2)]/[7 days x number of cycles with a non-zero dose].

The RDI of azacitidine is defined as the actual DI divided by the planned dose per day and expressed in %:

RDI (%) = 100× [DI (mg/ m^2/day]/[75 (mg/m^2/day)].

6.5.3.6. Dose Reductions and Interruptions due to Adverse Events

A dose reduction due to an AE is defined as a non-zero dose where reason for adjustment on the dosing CRF is AE.

The number and percentage of patients with at least one dose reduction as well as a breakdown of dose reductions (1 / 2 / ≥3) and/or the dosage administered will be summarized by treatment and by treatment arm.

An interruption due to an AE is defined as a 0 mg dose administered where reason for adjustment on the dosing CRF is AE. The number and percentage of patients with dose interruptions will be summarized by treatment and by treatment arm. Percentages will be calculated based on the total number of patients in the safety analysis set.

6.5.4. Concomitant Medications and Non-Drug Treatments

Concomitant medications and non-drug treatments received by patients during the study will be summarized for the safety analysis population by treatment arm.

Concomitant medications refer to all medications which started prior to first dose of study treatment and continued during the on-treatment period (see Section 3.5) as well as those started during the on-treatment period. Concomitant medications will be coded in the WHO Drug coding dictionary and will be tabulated by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term in descending order of frequency in the glasdegib arm. In case of equal frequency regarding drug class (respectively drug name), alphabetical

order will be used. A patient will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. Preferred Terms will be reported under each ATC class that it is included under within WHO Drug (no primary path is available in WHO Drug).

Concomitant non-drug treatments refer to all non-drug treatments administered during the on-treatment period. Non-drug treatments will be coded with the most current version of MedDRA and will be summarized by MedDRA's SOC and PT in descending order of frequency in the glasdegib arm. Patients will be counted only once per PT even if he/she received the same treatment multiple times.

Any medications or non-drug treatments which were only administered prior to treatment start will be listed but not summarized.

6.5.5. Prior and Subsequent Anti-Cancer Therapies/Procedures

Prior and subsequent Anti-Cancer Therapies and Procedures are defined as therapies entered on the 'Prior Cancer Therapy' and 'Follow-up Cancer Therapy' CRF pages, respectively. The number and percentage of patients within each category (medication therapy, radiation therapy, and surgeries) will be provided by treatment arm.

Medications will be coded using the WHO Drug coding dictionary and will be tabulated by preferred term in descending order of frequency on the glasdegib arm.

Analyses will be based on the FA set by treatment arm.

6.6. Safety Summaries and Analyses

Unless otherwise specified, summaries of AEs and other safety parameters will be based on the safety analysis population by treatment arm.

6.6.1. Adverse Events

All analyses will be based on treatment emergent events unless otherwise specified. Treatment emergent is defined in Section 3.5.1.

A high level summary of adverse events will include the number and percent of patients with:

- Any Adverse Event;
- Serious AE;
- Adverse Events with maximum CTCAE Grade 3-4;
- Grade 5 events;
- AEs leading to dose interruptions separately for each study drug, for chemotherapy, and for any study treatment;

- AEs leading to dose reductions separately for each study drug, for chemotherapy, and for any study treatment;

- AEs leading to withdrawal separately for each study drug, for chemotherapy, and for any study treatment.

Additionally, the number of events reported for each of the categories above will be provided. Each unique adverse event at the PT level in each treatment arm of the study for a patient is included in the count.

Seriousness, toxicity grade, action taken (interruption, reduction, and withdraw) are as reported by the investigator on the adverse event CRF.

Summaries by treatment arm, SOC and PT in decreasing frequency based on the frequencies observed in the glasdegib arm will be provided for:

- Treatment Emergent Events by Maximum Toxicity Grade (All Causality);

- Treatment Emergent Events by Maximum Toxicity Grade (Treatment Related);

- Serious Treatment Emergent Events by Maximum Toxicity Grade (All Causality);

- Serious Treatment Emergent Events by Maximum Toxicity Grade (Treatment Related).

An event will be considered treatment related if the investigator considered the event related to the study drug or this information is unknown.

The above four summaries will also be provided separately for (1) treatment emergent events prior to HSCT, (2) non-treatment emergent events during HSCT, and (3) treatment emergent events after HSCT.

The following summaries will be provided by treatment arm and PT (ie, summaries will not include SOC) in decreasing frequency based on the frequencies observed in the glasdegib arm for:

- Treatment Emergent Events (All Causality) Experienced by ≥10% of patients in either treatment arm;

- Treatment Emergent Events (All Causality) by Preferred Term and Maximum Toxicity Grade;

- Treatment Emergent Grade 3-5 Events (All Causality) by Preferred Term and Maximum Toxicity Grade;

- Treatment Emergent Adverse Events Leading to Dose Interruptions by Maximum Toxicity Grade (All Causality) separately for each study drug, for chemotherapy, and for any study treatment;

- Treatment Emergent Adverse Events Leading to Dose Reductions by Maximum Toxicity Grade (All Causality) separately for each study drug, for chemotherapy, and for any study treatment;

- Treatment Emergent Adverse Events Leading to Permanent Withdrawal by Maximum Toxicity Grade (All Causality) separately for each study drug, for chemotherapy, and for any study treatment;

- Serious Treatment Emergent Events (All Causality).

Each patient will be counted only once within each PT.

As described in Section 5.3.2, in case a patient has events with missing and non-missing toxicity grades, the maximum of the non-missing grade will be displayed. Missing grade will only be displayed in the event that only one event has been reported for a patient and the grade is missing.

6.6.1.1. Basic Results

For basic results disclosures in the US and EU the following additional summaries will be provided:

- Treatment Emergent Non Serious Adverse Events by SOC and PT in >5% of patients in either treatment arm;

- Treatment Emergent Serious Adverse Events by SOC and PT; and

- Fatal Adverse Events by SOC and PT.

Each of the above summaries will include a count of the number of patients with all causality events and the number of patients with treatment related events.

CCI

CCI [redacted]

6.6.1.3. Adverse Events of Special Interest

These analyses will be performed for treatment emergent AEs of special interest as specified in 3.5.1.

- **Time to AE Onset** (in days) is defined as the time from the date of the first dose to the onset date of the AE, regardless of grade. If a patient has multiple episodes of an AE, the date of the first occurrence is used. Time to AE onset (in days) will be calculated as (AE start date – first dose date +1). Time to onset is calculated for the subgroup of patients who had the specific AE.

- **Time to Grade 3 or 4 AE Onset** (in days) is defined similarly as time to AE onset for Grade 3 or 4 AEs.

- **Duration of AE** (in days) is defined as the cumulative duration across episodes of the AE, regardless of grade, where duration for each episode is the time from the AE start date to the AE end date. For one episode, duration (in days) = AE end date – AE start date + 1. If a patient has multiple episodes of an AE, cumulative duration across all episodes will be used adjusting for any overlap. If a patient has an AE that was ongoing at the time of analysis, the time is censored at the last available on treatment visit date. Duration is calculated for the subgroup of patients who had the specified AE.

Descriptive statistics will be presented for time to AE onset (days), time to Grade 3 or 4 AE onset, and duration of AEs for the subgroup of patients with the AE.

Person-year exposure analyses may also be considered in case of relevant difference in treatment duration between treatment arms.

6.6.2. Deaths

The frequency (number and percentage) of patients in the safety analysis set who died during the study and who died within 28 days after last dose of study treatment as well as the primary reason for death, will be tabulated based on information from the 'Notice of Death' CRF, by treatment arm.

The frequency (number and percentage) of patients in the safety analysis set who died within 30 days of first dose of study treatment will also be provided.

Date and cause of death will be provided in individual patient data listing together with selected dosing information (study treatment received, date of first/last administration, dose).

6.6.3. Laboratory Data

Laboratory results will be converted to International System of Units (Système International d'unités, SI) units which will be used for applying toxicity grades and for all summaries.

As described in Section 3.4, baseline will be defined as the last non-missing assessment performed on or prior to date of the first dose of study treatment (or prior to randomization for patients randomized but not dosed). If there are multiple assessments that meet the baseline definition on the same day without the ability to determine which was truly last, then the worst grade will be assigned as the baseline grade. Since a few CTCAE terms (including Hypo/Hypercalcemia and Activated Partial Thromboplastin) can be derived using several laboratory tests (analytes) refer to the section 2.3.7 of the "Pfizer Oncology CTCAE Grading Implementation Guidance for Laboratory Data" for determination of baseline CTCAE grade in this situation.

Laboratory results will be programmatically classified according to NCI-CTCAE version 4.03 grade. Non-numerical qualifiers will not be taken into consideration in the derivation of grade (eg, hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived). In summary statistics the number and percentage of patients corresponding to grades that only include non-quantitative criteria will be displayed as a blank or NA (not assessed) rather than 0. If there is any overlap between grade criteria (eg, CTCAE grading criteria for Creatinine Increased – a value can fall into one range based on comparison to ULN and another range based on comparison to baseline), the highest (worst) grade would be assigned to that record. Grade 5 is defined in the CTCAE criteria guidance as an event with an outcome of death. Since laboratory data does not collect an outcome, Grade 5 is not used when programmatically grading laboratory data.

Grade 0 or Outside Toxicity Reference (OTR) is not defined specifically in the CTCAE guidance. However, programmatically this is used as a category to represent those patients who did not meet any of the Grades 1 to 4 criteria. If the laboratory value is evaluable for CTCAE criteria grading (numeric value is present, valid units and ranges are present as required to allow conversion to standard units and grading), and does not qualify for any of the Grade 1-4 criteria for a given lab test, then the value is assigned as Grade 0 or OTR.

Abnormalities will be described using the worst grade post-baseline. When determining the maximum post-baseline grade for a given patient and CTCAE test, the maximum across all analytes and assessments contributing to that CTCAE test will be used. Several laboratory tests have bi-directional grading criteria defined so that both low (hypo) and high (hyper) values can be graded separately. Each criterion will be summarized separately. In the cases where a value is graded as a Grade 1, 2, 3, or 4 for one of the directions, that value will also be assigned as a Grade 0 for the opposite direction for that test. For example, a value meeting the criteria for Grade 3 Hypercalcemia will be classified as a Grade 0 Hypocalcemia. For CTCAE terms that can be derived using one of several laboratory tests, the maximum post-baseline grade for a given patient and CTCAE term will be the maximum across all possible laboratory tests.

For **WBC differential counts** (total neutrophil [including bands], lymphocyte, monocyte, eosinophil, and basophil counts), the absolute value will be used when reported by the lab. When only percentages are available (this is mainly applicable for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

Derived differential absolute count = (WBC count) × (Differential %value/100).

If the investigator reports both the absolute and % value for Neutrophils or Lymphocytes from the same laboratory sample date and patient, ONLY the absolute value will be graded. The % value will not be graded in this scenario.

If the % value is converted to the differential absolute count for grading and the LLN for the differential absolute count is not available (only LLN for % is available) then Grade 1 will be assigned if the following conditions are met:

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥800/mm^3.
- Neutrophil count decreased
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥1500/mm^3.

For **calcium**, CTCAE grading is based on Corrected Calcium and Ionized Calcium. Corrected Calcium is calculated from Albumin and Calcium as follows:

- Corrected Calcium (mmol/L) = Calcium (mmol/L) + 0.02 [40 - Albumin (g/L)].

Laboratory toxicities will be tabulated using descriptive statistics (number of patients and percentages):

- Summary of laboratory parameters by worst CTCAE grade post-baseline table will include number and percentage of patients with Grade 0, 1, 2, 3, 4, and Grade 1-4 laboratory abnormalities.
- Shift table will summarize baseline CTCAE grade versus the worst post-baseline CTCAE grade.
- Patients who developed toxicities of grade ≥3 will also be listed.

Additional laboratory results that are not part of NCI-CTCAE will be presented according to the following categories overall: below normal limit, within normal limits, and above normal limits. In the unlikely event that for a given patient, clinically significant abnormalities are noted in both directions (eg, > Upper Limit of Normal (ULN) and < Lower Limit of Normal (LLN)), then both abnormalities are counted.

Drug Induced Liver Toxicity

Alanine aminotransferase (ALT), aspartate aminotransferase (AST), Alkaline Phosphatase (ALP), and total bilirubin (TBILI) are used to assess possible drug induced liver toxicity. The ratios of test result over the ULN will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of patients with each of the following during the on-treatment period will be summarized by treatment arm:

- For patients with normal ALT, AST, and TBILI at baseline, (ALT or AST) > 3×ULN and TBILI > 2×ULN and ALP < 2×ULN or missing.
- For patients with ALT, AST, or TBILI above the ULN at baseline:
    - Preexisting AST or ALT baseline values above ULN: AST or ALT values >2 times the baseline values AND >3×ULN; or >8×ULN (whichever is smaller).
    - Preexisting values of TBILI above the normal range: TBILI level increased from baseline value by an amount of at least 1×ULN or if the value reaches >3×ULN (whichever is smaller).

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created, with different symbols for different treatment arms, by graphically displaying:

- peak serum ALT(/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT=3×ULN and total bilirubin=2×ULN.
- peak serum AST(/ULN) vs peak total bilirubin (/ULN) including reference lines at AST=3×ULN and total bilirubin=2×ULN.

In addition, a listing of all TBILI, ALT, AST and ALP values for patients with a post-baseline TBILI > 2×ULN, ALT > 3×ULN or AST > 3×ULN will be provided.

6.6.4. Vital Signs

Vital signs data includes weight, pulse, systolic blood pressure, and diastolic blood pressure. Measurements were only to be provided once per timepoint. If multiple assessments are provided per timepoint, the maximum value will be used for reporting.

Vital signs data will be summarized using simple descriptive statistics (mean, standard deviation, median, quartiles, minimum, and maximum) of actual values and change from baseline for each nominal visit over time (ie, unscheduled assessments will be excluded). The total number of patients for change from baseline will include all patients in the treatment arm who have both a baseline value and a value at the nominal visit. Baseline will be selected as defined in Section 3.4.

The number and percent of patients in each of the following minimum and maximum blood pressure, body weight, and pulse categories will be presented:

- Increase in Systolic Blood Pressure ≥40 mmHg.
- Decrease in Systolic Blood Pressure ≥40 mmHg.
- Decrease in Systolic Blood Pressure ≥60 mmHg.
- Increase in Diastolic Blood Pressure ≥20 mmHg.
- Decrease in Diastolic Blood Pressure ≥20 mmHg.
- Decrease in Diastolic Blood Pressure ≥40 mmHg.
- Decrease in Body Weight ≥10%.
- Maximum Pulse Rate >120 bpm.
- Minimum Pulse Rate <50 bpm.
- Maximum increase in pulse rate ≥30 bpm.
- Maximum decrease in pulse rate ≥30 bpm.

All assessments, including unscheduled assessments will be considered. A patient can be included in multiple categories if different criteria are met at different timepoints.

6.6.5. Electrocardiogram

Triplicate ECGs assessments are required at each specified nominal visit and time-point as specified in the Schedule of Activities in the Protocol (Table 2 for Intensive and Table 4 for Non-intensive). A mean value is calculated for any replicate measurements having the same nominal visit. The mean measurement is reported.

QTcF and QTcB (both Fridericia and Bazett's correction) will be programmatically derived using the following formula:

$$QTcF(msec) = QT(msec)/(60/HR[bpm])^{1/3}$$

$$QTcB(msec) = QT(msec)/(60/HR[bpm])^{1/2}$$

QT, RR, HR, PR, QRS, QTcF, and QTcB will be summarized using simple descriptive statistics (mean, standard deviation, median, quartiles, minimum, and maximum) of actual values and change from baseline for each nominal visit over time (ie, unscheduled assessments will be excluded). The total number of patients for change from baseline will include all patients in the treatment arm who have both a baseline and a value at the nominal visit. Baseline will be selected as defined in Section 3.4.

The mean absolute value will be presented with two-sided 95% confidence intervals and the baseline adjusted mean QT, RR, HR, PR, QRS, QTcF and QTcB will be presented with two-sided 90% confidence intervals.

Additionally QTcF and QTcB will be summarized by maximum post-baseline values using the following categories.

- ≤450 msec;
- >450 msec but 480 msec;
- >480 msec but ≤500 msec;
- >500 msec.

Unscheduled assessments will be utilized in addition to planned assessments.

And maximum increases from baseline in QTcF and QTcB (including scheduled and unscheduled assessments) will be also summarized based on the following categories:

- Change ≥60 msec;
- Change ≥30 msec but <60 msec;
- Change <30 msec.

In the event that neither Fridericia's nor Bazett's correction adequately adjusts for heart rate, an additional correction such as a population or patient-specific baseline correction could be used and should be fully justified.

For PR and QRS maximum increases from baseline, the following categories will be applied:

- PR changes from baseline ≥25% and absolute values >200 msec;
- QRS changes from baseline ≥25% and absolute values >110 msec.

Shift tables will be provided for baseline versus worst on study QTc (one or more correction method will be used) using categories ≤450 msec, >450 msec but ≤480 msec, >480 msec but ≤500 msec, and >500 msec. Tables of ECG abnormality at baseline (yes, no, not done: (n, %)) will also be provided. Subjects experiencing abnormal and clinically-relevant morphological ECG changes will be summarized (including frequency and percentage).

If more than one ECG is measured at a nominal time post-dose (eg, triplicate ECGs within 2-4 minutes), the mean will be used to represent a single observation per patient and time post-dose. If any of the three individual ECGs results in a QTc >500 msec and the mean is not >500 msec, then that patient's data will be described in the safety section in the study report in order to place the >500 msec value in appropriate clinical context. On the other hand, such individual >500 msec value within a triplicate will not be included in the categorical analysis unless the average from that triplicate is also >500 msec. Data listings will contain the means from a triplicate as well as the parameters from each of the three ECGs. Note that using the mean value may result in a patient having a measurement that is not represented by an actual ECG.

6.6.6. Physical Examination

Physical examination findings will only be listed.

6.6.7. Left Ventricular Ejection Fraction (LVEF)

LVEF% will be summarized using simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each nominal visit over time. In addition, a shift table will be provided for baseline versus worst on-treatment LVEF% per NCI CTCAE version 4.03.

A patient will be included in the categories above if any post-baseline assessment (including unscheduled assessments) meet the criteria; however only post-baseline assessments which use the same method of assessment (ECHO or MUGA) as baseline will be considered.

6.6.8. Performance Status

The ECOG shift from baseline to the highest score during the post-baseline period will be summarized by treatment arm.

CCI

[redacted]

7. INTERIM ANALYSES

7.1. Introduction

The goals of the interim analyses are to allow early stopping of the study for futility or efficacy. The interim analysis of OS will be performed as described in Section 5.1.1 and 5.1.2 using the methodology described in Section 6.2.1 for OS.

The interim analysis will be performed by an independent statistician. Unblinded results from the interim analysis will not be communicated to the Sponsor's clinical team or to any party involved in the study conduct (apart from the independent statistician and E-DMC members) until the E-DMC has determined that (i) OS analysis has crossed the prespecified boundary (except for the efficacy boundary at the first planned interim analysis for the intensive chemotherapy study when there is no plan to stop for efficacy regardless of whether the boundary is crossed) and the Sponsor has decided to terminate the study or (ii) the study needs to be terminated due to any other cause, including safety reasons. Further details will be described in the E-DMC charter.

The final OS analysis will be performed by the Sponsor's clinical team.

7.2. Interim Analyses and Summaries

At each analysis timepoint, the critical boundaries for the group sequential test will be derived from the predefined spending function(s) as described in Section 5.1. The calculations of boundaries will be performed using EAST. For the intensive chemotherapy study, there will not be any efficacy stopping at the first planned interim analysis even if the efficacy stopping boundary is crossed.

Let $l(t_1)$ and $u(t_F)$ denote the lower critical boundaries for futility at the interim analysis and the upper critical boundaries for efficacy at the final analysis based on the test statistics Z_1 and Z_F, respectively.

The critical values $l(t_1)$ for the interim analysis of OS are determined such that

$$P_a(Z_1 \le l(t_1)) = \beta(t_1)$$

where P_a denote the probabilities under the alternative hypothesis and $\beta(t_1)$ denote the β spent, based on the predefined spending functions at information fraction t_1 (t_1 is calculated as the ratio of the number of deaths observed at the time of the data cutoff for the interim analysis and the total number of deaths targeted for the final analysis).

The boundary for the final efficacy analysis is derived such that

$$P_0(Z_F \ge u(t_F)) = 0.025$$

8. REFERENCES

1. Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017; **129**(4): 424-47.

2. Brookmeyer R and Crowley J. A confidence interval for the median survival time. Biometrics 38:29-41, 1982.

3. Kalbfleisch JD, Prentice, RL. Statistical Analysis of Failure Time Data, 2nd Edition. Hoboken, Wiley Interscience.

4. Schoenfeld, D. A. (1982), "Partial Residuals for the Proportional Hazards Regression Model," Biometrika, 69, 239–241.

5. Therneau, T. M. and Grambsch, P. M. (2000). Modeling Survival Data: Extending the Cox Model, Springer-Verlag, New York.

6. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4): 183–9.

7. Fayers P et al. EORTC QLQ-C30 Scoring Manual. 2001. EORTC, Brussels.

8. The EuroQol Group. EQ-5D-5L: User Guide. V 2.1, April 2015. EuroQol Research Foundation, Rotterdam.

9. Osoba D et al, Interpreting the significance of changes in health-related quality of life scores. 3. Cin Oncology. 1998. 16:139-144.

10. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. 1997. Biometrics, 53, 983-997.

11. Santner TJ and Snell MK. Small-Smple Confidence Interval for $p_1 - p_2$ and p_1/ p_2 in 2x2 Contingency Tables. Journal of the American Statistical Association 75: 386-394, 1980.

12. Fernandez HF, Sun Z, Yao X, et al. Anthracycline dose intensification in acute myeloid leukemia. *The New England journal of medicine* 2009; **361**(13): 1249-59.

13. Lowenberg B, Ossenkoppele GJ, van Putten W, et al. High-dose daunorubicin in older patients with acute myeloid leukemia. *The New England journal of medicine* 2009; **361**(13): 1235-48.

14. Dombret H, Seymour JF, Butrym A, et al. International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. *Blood* 2015; **126**(3): 291-9.

15. Schemper M and Smith TL, A Note on Quantifying Follow-up in Studies of Failure Time, *Controlled Clinical Trials* 1996, **17**:343-346.

16. Morgan TM, Analysis of uration of rResponse: A Problem of Oncology Trials, *Controlled Clinical Trials* 1988, **9**:11-18.

17. Gray RJ, A class of k-sample tests for comparing the cumulative incidence of a competing risk, *Annals of Statistics* 1988, **16**: 1141-1154.

9. APPENDIX

9.1. Note on Defining CIR for "Fatigue" Single-item from the MDASI AML/MDS Questionnaire

(Note this description is for the cohort of intensive patients. The analysis for non-intensive patients can easily be inferred by extending Week 8 to Week 12 throughout).

FDA document "*Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*" gives following definition for Clinically Important Responder (CIR) Definition:

> *"Responder definition — A score change in a measure, experienced by an individual patient over a predetermined time period that has been demonstrated in the target population to have a significant treatment benefit."*

In the guidance the following algorithm is proposed to define responder definition:
*"…anchor-based approach to defining responders makes use of patient ratings of change administered at different periods of time or upon exit from a clinical trial. These numerical ratings range from **worse** to **the same** and **better**. The difference in the PRO score for persons who rate their condition the same and better or worse can be used to define responders to treatment. Patient ratings of change are less useful as anchors when patients are not blinded to treatment assignment…."*

> First ,we do not have simple numerical rating scale with just 3 categories – that is, "*numerical ratings range from **worse** to **the same** and **better***." PGIS is used to create a simplified numerical rating with 3 categories ***worse, the same,** and **better.***
>
> Specifically, we create a new anchor Subject Global Impression of Change using PGIS (SGIC-S) by using the following algorithm (note that Y in below represents "Week 1", "Week 2",… "Week 8"):

a. SGIC-S = -1, if PGIS at Week Y minus PGIS at baseline (day 1) is positive (this means worse);

b. SGIC-S = 0, if PGIS at Week Y minus PGIS at baseline (day 1) is zero (ie, no change, this means the same);

c. SGIC-S = 1, if PGIS at Week Y minus PGIS at baseline (day 1) is negative (this means better).

A repeated measures model will be used to determine CIR. Figure 1 represents the SAS implementation of the model. We use the *change* in the "Fatigue" single-item (from baseline) as the outcome (variable changeFatigue) and SGIC-S score as the anchor variable (note that variable SGICS has only 3 categories ***worse** [value of -1], **the same** [value of 0], and **better** [value of 1]*). It is important to note that we treat SGIC-S in this model as a continuous variable – by doing so we impose a linear relationship between change in a "Fatigue" single-item and SGIC-S. As a sensitivity analysis we may also perform modeling using SGIC-S as a categorical variable – this approach does not impose any functional

relationship between outcome and anchor (see figure 2 for SAS implementation). In this case, the larger difference between two contiguous categories (no change vs. worse, or better vs. no change) will be used as the CIR. We should note that all available post-baseline data will be used (starting with Week 1 and up to Week 8 from the cohort of intensive patients).

Figure 1. SAS Proc Mixed code (SGIC-S as a continuous anchor)

```
Proc mixed data=_mixed_;

   Class PID Week ;
   Model changeFatigue = SGICS / ddfm=kr s;

   Repeated Week / Type=SP(POW)(cVisit)  Subject=PID;

   Estimate "CIR (One Category Change)"  SGICS 1 /cl;

   Estimate "worse    SGIC=-1" Intercept 1 SGICS -1 /cl;
   Estimate "the same SGIC=0 " Intercept 1 SGICS 0 /cl;
   Estimate "better   SGIC=1 " Intercept 1 SGICS 1 /cl;

   Run;
```

Figure 2. SAS Proc Mixed code (SGIC as a categorical anchor)

```
Title "Experience Domain Score";
Proc mixed data=_mixed_;

   Class PID Week SGICS;
   Model changeFatigue = SGICS / ddfm=kr s;
   Repeated Week / Type=SP(POW)(cVisit)  Subject=PID;
   LSMeans  SGICS / diff cl;

Run;
```

In addition, sensitivity analyses the above approach is augmented by using PGIC. Specifically, we create another anchor Subject Global Impression of Change using PGIC (SGIC-C) by using following algorithm (note that Y in below represents "Week 1", "Week 2",… "Week 8"):

a. SGIC-C = -1, if PGIC at Week Y represents worsening (PGIC scores of 5, 6, and 7; this means ***worse***);

b. SGIC-C = 0, if PGIC at Week Y is 4 (ie, no change, this means ***the same***);

c. SGIC-C = 1, if PGIC at Week Y represents improvement (PGIC scores of 1, 2, and 3; this means ***better***).

The same models represented by figures 1 and 2 will be used (the variable SGICS will be replaced by variable SGICC representing SGIC-C).