NOVARTIS

Clinical Development

MBG453/Sabatolimab

CMBG453B12301 / NCT04266301

A randomized, double-blind, placebo-controlled phase III multi-center study of azacitidine with or without MBG453 for the treatment of patients with intermediate, high or very high risk myelodysplastic syndrome (MDS) as per IPSS-R, or Chronic Myelomonocytic Leukemia-2 (CMML-2)

Statistical Analysis Plan (SAP)

Author:	Trial Statistician, [redacted]
Document type:	SAP Documentation
Document status:	Final version 3.0 - Amendment 2
Release date:	18-Sep-2023
Number of pages:	47

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
27-May-20	FPFV	Creation of version 1.0	N/A - First version (based on protocol from 15-Nov-19)	N/A
27-Jun-2022	Prior to efficacy IA DBL	Creation of version 2.0	Amendment 1 (based on protocol amendment v03 from 27Jun2022)	-Typos corrections and clarifications. - Naming change of study drug to investigational drug -Section 2.3.2 Addition of a rationale for collecting race. -Section 2.4.1 Study treatment / compliance and appendix: update of the dose intensity and RDI calculation. -Section 2.4.2 Clarifications added for transfusion analyses and for prior anti-neoplastic therapies. -Section 2.5.3: Added OS censoring reason. -Section 2.5.4: Supportive analysis added for OS considering the hypothetical strategy for intercurrent events (anti-neoplastic therapy and HSCT). -Section 2.6.1.1 Estimand definition and censoring rules were added for the analysis of time to definitive deterioration of fatigue. -Section 2.6.1.5 Estimand definitions were added for fatigue, physical and emotional function analyses. -Section 2.7.1 Addition of CR +PR for response rate analysis. Clarification on response rate subgroup analyses was added. Censoring reasons updated to take into account discontinuation from follow-up/study due to subject/guardian/physician decision. Modification of the definition of transfusion independence/dependence. -Section 2.7.2 Additional AE tables, especially one for COVID-19 related AEs and two for AEs/SAEs for safety follow up period (day 31-150 after last dose of investigational drug). -Section 2.7.3 Addition of listings for sample ADA status and subject ADA status. Note added to clarify that details on ADA analyses will be provided in a separate document. [redacted] - Section 2.8.2 Changes from baseline for EORTC QLQ-C30 will also be displayed graphically. [redacted]

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
				-Section 2.9 and 2.10 and in other SAP sections: modifications as per protocol amendment v03. The main purpose of this amendment was to adjust the group-sequential statistical plan based on independent information that became available from the final progression free survival (PFS) analysis of the CMBG453B12201 study, which was conducted in a similar patient population and study treatment. Based on these data, a delayed treatment effect is considered for this study which justifies an increase in overall survival (OS) events to retain statistical power for the primary analysis at 85%. In order to achieve a statistical power of 85%, based on assumption of 5 months delayed treatment effect and followed by an effect of the same magnitude as assumed in the protocol version 00 (i.e. HR=0.6), the number of OS events for the primary analysis is increased from 180 to approximately 282 events. With increase from 180 to 282 OS events, the information fraction corresponding to the 135 events planned for efficacy IA changed from 75% to 48% and thus hazard ratio threshold and alpha had to be changed. Since the futility analysis was already conducted under the design of protocol version 00 (40% information fraction out of 180 events), no changes were made to the threshold for declaring futility or its calculation. -Appendix updated for calculations of date of last exposure to investigational drug and AZA.
15-Sep-2023	Prior to primary analysis DBL and unblinding	Creation of version 3.0	Amendment 2 (based on protocol amendment v03 from 27Jun2022)	-Typos corrections and clarifications. -Section 2.2 Addition of specification of immunogenicity analysis sets. -Section 2.4.5.2 Addition of sensitivity estimand, in case of substantial misclassification in the randomization strata. -Section 2.5.1.2 Addition of permutation test for RBC transfusion-free intervals in case the negative binomial regression model fails to converge. -Section 2.5.1.5 Note added to clarify for key secondary endpoints, the second improvement of fatigue, physical functioning and emotional functioning can be at any time. -Section 2.5.1.6 Addition of supplementary estimand by including all RBC transfusion records regardless of reason given; Addition of supplementary estimand with alternative thresholds; Addition of 'while-on-treatment' supplementary estimand. -Section 2.6.1 Note added to clarify that for LFS the same censoring rules as PFS will be applied. [redacted]

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
				[redacted] -Section 2.6.2: Update vital sign section to state that all patients will be analyzed regardless of their baseline value. -Section 2.6.4 Addition that also absolute values for PRO will be summarized descriptively. [redacted] -Section 2.7.2: Note added to clarify that the anchor-based analysis will be provided in a separate report. -Appendix: Update derivation of duration of exposure for investigational drug, AZA and study treatment. Addition of IPSS-M classification derivation rules.

Table of contents

List of abbreviations

ADA	Anti-drug Antibody
AE	adverse event
AESI	Adverse events of special interest
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BLQ	Below the Limit of Quantitation
BM	Bone marrow
BMI	Body mass index
BOR	Best Overall Response
BP	Blood pressure
BSA	Body surface area
CI	Confidence interval
CR	Complete Remission
CRO	Contract research organization
CSR	Clinical study report
CTC	Common Terminology Criteria
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DMC	Data Monitoring Committee
DMS	Document Management System
eCRF	Electronic Case Report/Record Form
eCRS	Electronic case retrieval strategy
ECOG	Eastern Cooperative Oncology Group
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer - Quality of Life Questionnaire
EOT	End of treatment
EQ-5D-5L	EuroQol Group - standardized measure of health status questionnaire
FAS	Full Analysis Set
HLT	High level terms
AZA	Hypomethylating agent
HR	Hazard ratio
HSCT	Hematopoietic stem cell transplantation
IV	Intravenous
IA	Interim Analysis
IG	Immunogenicity
IPSS-R	International Prognostic Scoring System
IRT	Interactive Response Technology
IWG	International Working Group
mCR	marrow Complete Remission
MDS	Myelodysplastic syndromes

MedDRA	Medical dictionary for regulatory activities
mg	milligram(s)
[redacted]	[redacted]
NCI	National Cancer Institute
NGS	Next-Generation Sequencing
NMQ	Novartis MedDRA queries
OS	Overall survival
PAS	Pharmacokinetic Analysis Set
PD	Pharmacodynamics
PFS	Progression free survival
[redacted]	[redacted]
[redacted]	[redacted]
PK	pharmacokinetic(s)
PR	Partial remission
PRO	Patient Reported Outcome
PS	Performance status
PT	Preferred term
Q4W	Every 4 weeks
QoL	Quality of life
RBC	red blood cell(s)
SAE	serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
SMQ	standardized MedDRA queries
SOC	Standard of Care
sTIM-3	Soluble T-cell immunoglobulin domain and mucin domain-3
TFL	Tables Figures Listings
TIM-3	T-cell immunoglobulin domain and mucin domain-3
VAS	Visual Analog Scale
VAF	Variant of Allele Frequency
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes the planned analyses for the primary Clinical Study Report (CSR) of the study CMBG453B12301, a randomized, double-blind, placebo-controlled phase III multi-center study of azacitidine with or without MBG453 for the treatment of patients with intermediate, high or very high risk myelodysplastic syndrome (MDS) as per IPSS-R, or Chronic Myelomonocytic Leukemia-2 (CMML-2).

As specified in the Section 12.7 of the study protocol, the futility interim analysis will be performed when approximately 72 OS events have been observed. In case the study is terminated for futility at that time, the futility interim analysis will constitute the basis of the primary CSR. Otherwise, the next OS analysis (at the efficacy interim or at the primary analysis), which meet statistical significance, will constitute the basis of primary CSR. Updated analyses after the primary CSR will be conducted and reported as needed. This SAP will serve as the basis for those analyses as well, however, a separate selection of tables, figures and listings (TFL) might be done.

The content of this SAP is based on the CMBG453B12301 protocol including amendment (27-Jun-2022). All decisions regarding the analysis, as defined in the SAP document, were made prior to database lock and unblinding of the study data.

1.1 Study design

This Phase III is a randomized, double-blind, placebo-controlled, multi-center phase III study of MBG453 or placebo added to azacitidine for the treatment of subjects with intermediate, high or very high risk MDS as per IPSS-R or with CMML-2.

Subjects will be randomized in a 1:1 ratio as described in Figure 1-1 to one of the following:

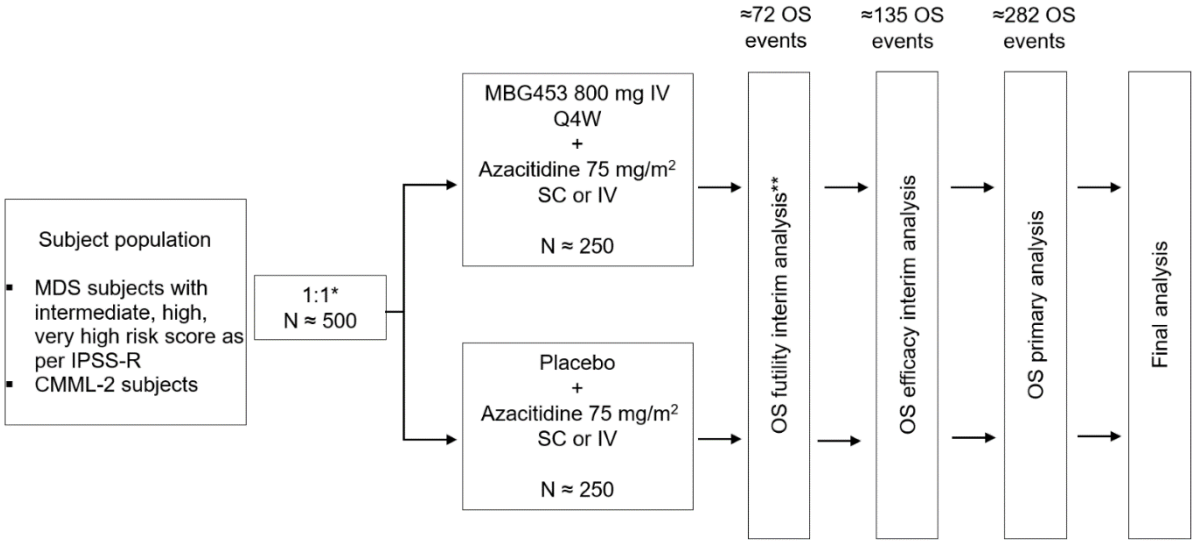
- MBG453 + azacitidine or
- Placebo + azacitidine

The randomization will be stratified into 4 groups based on the status at study entry:

- Intermediate risk MDS,
- High risk MDS,
- Very high risk MDS,
- CMML-2

Study treatment consists of cycles of MBG453 or placebo 800 mg IV Q4W administered on Day 8 of each cycle in combination with azacitidine administered to the subjects on days 1 to 7 (or on days 1 to 5 and days 8 and 9) of each cycle until treatment discontinuation as described in the protocol. The planned duration of a cycle is 28 days. Crossover between treatment arms is not permitted at any time during the study.

Figure 1-1 Study design



* The randomization will be stratified into 4 groups: intermediate risk MDS, high risk MDS, very high risk MDS and CMML-2.
** At futility analysis, the available data from phase II study MBG453B12201 will be used to inform the futility assessment as described in protocol section 16.4.

After the end of study treatment, all subjects must be followed for adverse events (AEs) for 30 days following the last dose of azacitidine, or 150 days following the last dose of MBG453 or placebo, whichever is later. In addition, all subjects who discontinued study treatment will enter a long-term follow-up (for efficacy and/or survival status) for up to 5 years from the last subject randomized. Subjects who are scheduled for hematopoietic stem-cell transplant (HSCT) or intensive chemotherapy at any time during the course of the study will be discontinued from study treatment and will enter a long term follow-up (for efficacy and/or for survival status) for up to 5 years from the last subject randomized.

Data Monitoring Committee (DMC)

This study will include a DMC which will function independently of all other individuals associated with the conduct of this clinical study, including the site investigators participating in the study. The DMC will regularly assess safety data and recommend to the sponsor whether to continue, modify, or terminate the study. The DMC will review also overall survival and relevant efficacy data at the time of the futility interim analysis and efficacy interim analysis.

For any details, please refer to the DMC Charter as well as the separate planning documents for the DMC analyses. However, the DMC analyses will use this SAP as basis with regard to definition of endpoints, analyses and statistical testing plan.

1.2 Study objectives and endpoints

The following Table 1-1 (which is a copy of the Table 2-1 from the study protocol) outlines the primary, secondary [redacted] objectives and belonging endpoints. Further details are given in the statistical methods section of this SAP.

Table 1-1 Objectives and related endpoints

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• To compare overall survival (OS) in the MBG453 plus azacitidine arm versus placebo plus azacitidine arm	• OS is the time from randomization until death due to any cause. If the subject is not known to have died, then OS will be censored at the latest date the subject was known to be alive (on or before cut-off date).
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**
• **Key Secondary Objective 1**: To compare time to definitive deterioration of fatigue in the MBG453 plus azacitidine arm versus placebo plus azacitidine arm	• Time from randomization to at least 3 points worsening from baseline in FACIT-fatigue scores with no subsequent improvement above this threshold or death due to any cause, whichever occurs first
• **Key Secondary Objective 2**: To compare RBC transfusion-free intervals in the MBG453 plus azacitidine arm versus placebo plus azacitidine arm	• Cumulative time of intervals with no evidence of RBC transfusion for at least 8 weeks at any point after randomization
• **Key Secondary Objective 3:** To compare improvement of fatigue in the MBG453 plus azacitidine arm versus placebo plus azacitidine arm	• Percent of subjects with at least 3 point confirmed improvement from baseline in FACIT-fatigue scores
• **Key Secondary Objective 4:** To compare improvement of physical functioning in the MBG453 plus azacitidine arm versus placebo plus azacitidine arm	• Percent of subjects with at least 10 point confirmed improvement from baseline in physical functioning using EORTC QLQ-C30
• **Key Secondary Objective 5:** To compare improvement of emotional functioning in the MBG453 plus azacitidine arm versus placebo plus azacitidine arm	• Percent of subjects with at least 10 point confirmed improvement from baseline in emotional functioning using EORTC QLQ-C30
• To assess response rate in each treatment arm	• - Percentage of CR/mCR/PR/HI according to IWG-MDS as per investigator assessment - Percentage of SD according to IWG-MDS as per investigator assessment
• To assess PFS in each treatment arm	• Time from randomization to disease progression (including transformation to acute leukemia per WHO 2016 classification), relapse from complete remission (CR) according to IWG-MDS or death due to any cause, whichever occurs first, as per investigator assessment
• To assess Leukemia-free survival in each treatment arm	• Time from randomization to ≥ 20% blasts in bone marrow/peripheral blood (per WHO 2016

Objective(s)	Endpoint(s)
	classification) or diagnosis of extramedullary acute leukemia, or death due to any cause
• To assess the safety profile of MBG453 when given in combination with azacitidine	• Incidence and severity of AEs and SAEs, changes in laboratory values and vital signs (per CTCAE version 5)
• To assess the improvement in RBC/platelets transfusion independence in each treatment arm	• Number and percent of transfusion dependent subjects at baseline who become RBC/platelets transfusion independent after randomization as per IWG-MDS criteria
• To characterize the pharmacokinetics of MBG453	• Serum concentrations and pharmacokinetic parameters for MBG453
• To evaluate immunogenicity of MBG453	• Anti-drug Antibody (ADA) prevalence at baseline and ADA incidence on-treatment
• To assess overall quality of life in each treatment arm	• - Change from baseline in EQ-5D-5L scores and VAS scores over time - Change from baseline to C12D1 of Global Health Status/QoL scores using EORTC QLQ-C30

[redacted]

2 Statistical methods

2.1 Data analysis

The primary analysis will be performed by Novartis. However, the DMC analyses will be performed by an Independent Statistician and Independent Programmer at a CRO. For details, refer to the DMC Charter and the separate DMC planning documents.

SAS version 9.4 or later R 3.4.3 or later software will be used to perform all data analyses and to generate tables, figures and listings.

Data included in the analysis / data cut-off handling

For each of the analyses a data cut-off date will be established after the targeted number of events for the planned futility interim analysis, efficacy interim analysis and primary analysis has been documented.

For each analysis time point, analyses will be performed using all data collected in the database up to the data cut-off date. All data with an assessment date or event start date (e.g. laboratory assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.

General analysis conventions

Data will be summarized by treatment arm: demographics and other baseline characteristics as well as efficacy data and anti-neoplastic therapies / HSCT for the Full Analysis Set, safety and any other data (unless specified otherwise) for the Safety Set.

Qualitative data (e.g., gender, race) will be summarized by means of contingency tables; a missing category will be included as applicable. Percentages will be calculated using the number of subjects in the relevant population or subgroup as the denominator.

Quantitative data (e.g., age, body weight) will be summarized by appropriate descriptive statistics (e.g. mean, standard deviation, median, 25^{th} and 75^{th} percentiles, minimum, and maximum).

Study treatment and investigational drug

MBG453 matching placebo will be referred to as "placebo" and azacitidine as "AZA" throughout this document and in the analyses. The treatment arms are MBG453 + AZA and placebo + AZA.

Study treatment refers to the combination of MBG453 or placebo with AZA.

Investigational drug refers to the individual components: MBG453 or placebo.

2.2 Analysis sets

Full Analysis Set

The Full Analysis Set (FAS) comprises all subjects to whom study treatment has been assigned by randomization. According to the intent to treat principle, subjects will be analyzed according to the treatment and stratum they have been assigned to during the randomization procedure.

Safety Set

The Safety Set includes all subjects who received at least 1 dose of any component of the study treatment (MBG453 + AZA or placebo + AZA). Subjects will be analyzed according to the study treatment they received, either MBG453 + AZA or placebo + AZA. If the subject never received the investigational drug (i.e. MBG453 or placebo) and took at least 1 dose of AZA, subjects will be analyzed in the placebo + AZA treatment arm.

Pharmacokinetic Analysis Set

The Pharmacokinetic Analysis Set (PAS) includes all subjects in the Safety Set, who had at least 1 evaluable PK concentration.

For a concentration to be evaluable:

- Dosing information must be properly documented (data and time of administration)
- For post-dose samples: planned dose of MBG453/placebo must be taken prior to sampling
- For pre-dose samples: the sample is collected before the next dose administration

Immunogenicity (IG) analysis sets

The Immunogenicity prevalence set includes all subjects in the Safety set with a non-missing baseline ADA sample or at least one non-missing post-baseline ADA sample.

The Immunogenicity incidence set includes all subjects in the Immunogenicity prevalence set with a non-missing baseline ADA sample and at least one non-missing post-baseline ADA sample.

Subject Classification

Subjects may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific subject classification rules defined in Table 2-1.

Table 2-1 Subject classification based on protocol deviations and non-protocol deviations criteria

Analysis set	Protocol deviations leading to exclusion	Non protocol deviation leading to exclusion
FAS	No written informed consent	None
Safety Set	No written informed consent	No dose of any component of study treatment
PAS	No written informed consent	See definition of PAS
IG Set	No written informed consent	See definition of IG analysis Set

Withdrawal of Informed Consent

Any data collected in the clinical database after a subject withdraws informed consent from all further participation in the study will not be included in the analyses. The date on which a subject withdraws consent is recorded in the eCRF.

Additional data for which there is a separate informed consent, e.g. biological sample etc., collected in the clinical database without having obtained that consent or after withdrawal of consent will not be included in the analyses.

2.2.1 Subgroup of interest

The main subgroups of interest are the stratification factor as well as new disease classification that became available for which the primary efficacy endpoint of OS will be analyzed:

- Randomization stratification factor (intermediate risk MDS, high risk MDS, very high risk MDS, CMML-2)
- <10% vs >=10% bone marrow blasts (also in light of new ELN2022 criteria (Doehner et al, EHA 2022))

[redacted]

2.3 Subject disposition, demographics and other baseline characteristics

2.3.1 Subject disposition

Number (%) of subjects screened will be summarized by country and center. In addition, the number (%) of subjects randomized will be summarized by country, center and treatment arm. For subjects who are screen failures, the reasons for not completing screening will be summarized based on "Screening Phase Disposition" eCRF.

The number (%) of subjects in the FAS who started treatment, are still on treatment, who entered and discontinued post-treatment follow-up and the study after survival follow-up will be summarized together with the respective reasons for treatment/post-treatment follow-up/end of study discontinuation.

All disposition information will be listed.

Protocol deviations

The number (%) of subjects in the FAS with any protocol deviation will be tabulated by deviation category. COVID-19 related protocol deviations will be summarized separately. All protocol deviations will be listed.

Analysis sets

The number (%) of subjects in each analysis set will be summarized by treatment arm and randomization stratum for the FAS. A listing will be provided displaying all subjects excluded from analysis sets.

2.3.2 Demographic and other baseline characteristics

Demographic and other baseline data including disease characteristics will be summarized descriptively and listed.

MBG453 is a novel compound (tested in a novel combination) and the collection of information on race is required to perform subgroup analysis. The aim is to detect potential signal in safety and efficacy.

BMI (kg/m2) at baseline will be calculated as weight[kg] / (height[m]^2) using weight at baseline and height at screening. Body Surface Area (BSA) is based on the Mosteller formula described in Section 2.4.1.

Details on MDS and CMML-2 diagnosis (initial diagnosis, WHO classification, current disease status (de novo or secondary) and cytogenetic abnormalities) will be tabulated and time since diagnosis summarized.

Data from the IRT system at randomization will be summarized also: randomization stratification factor including the IPSS-R risk category components (blasts in BM, number of cytopenias and cytogenetic abnormalities).

Medical history

Relevant medical histories and current medical conditions at baseline will be summarized by system organ class and preferred term. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable outputs.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

Duration of exposure

The duration of exposure (in months) will be summarized for study treatment and for each investigational drug individually (MBG453 and placebo) and for AZA based on summary statistics and categorical analyses (e.g. exposure <1 month, at least 1 month, at least 2 months

etc.). Details on start and end dates used for derivations are outlined in the Appendix of this SAP.

Dose intensity

The actual dose intensity (computed as the ratio of actual cumulative dose received and duration in days with at least one component of the combination) and the relative dose intensity (computed as the ratio of actual dose intensity and planned dose intensity) will be summarized for each study treatment component by descriptive statistics.

For MBG453/placebo, the actual cumulative dose in mg is the sum of "dose administered" from the eCRF of all cycles during the exposure of MBG453/placebo.

For AZA, the actual dose in mg/m^2 in each cycle is the "dose administered" in mg during that cycle divided by the body surface area (BSA) at the beginning of the cycle using the weight measured before the infusion at that cycle. The actual cumulative dose in mg/m^2 is then the sum of all cycles. The following formula is used for BSA:

$$\text{BSA } (m^2) = \sqrt{\text{Weight (kg)} * \text{Height at screening (cm)}/3600} \text{ (Mosteller formula)}$$

The duration considered for the derivation of the dose intensity and the relative dose intensity will be derived from the start date of study treatment to the end of the last cycle initiated irrespective of date of death, last contact date for withdraw consent and cut-off date: the last exposure to study treatment (combination) will be the planned end date (Day 28) of the last cycle initiated with MBG453 and/or AZA, whichever is the latest. Details are provided in the Appendix.

The relative dose intensity is then comparing the actual dose intensity during subjects' exposure with the protocol planned dose of 800 mg IV Q4W for MBG453/placebo and 75 mg/m^2 for azacitidine, e.g. if a subject received 600 mg Q4W on average throughout the study, the relative dose intensity for this subject is 0.75.

Details on the duration from the first to last cycle initiated for the derivation of the dose intensity and the relative dose intensity are provided in the Appendix of this SAP.

Dose reductions, interruptions or permanent discontinuations

The number (%) of subjects with any dose changes (incl. reductions, interruptions, or permanent discontinuations) and the reasons (e.g. AE, dosing error, dispensing error, physician decision) will be taken from the 'Study Treatment eCRF' and summarized by investigational drug. The total duration of interruptions by subject will be summarized for the study population by time intervals, e.g. <1week, ≥1-<2 weeks, ≥2-<3 week etc. (these time intervals may be adjusted depending on the observed data).

2.4.2 Concomitant and post-treatment therapies / HSCT

Prior and concomitant medications/therapies

Concomitant medications and significant non-drug therapies prior to and after the start of the study treatment will be summarized according to the Anatomical Therapeutic Chemical (ATC) classification system.

Prior anti-neoplastic therapies

Prior anti-neoplastic medications will be summarized. Medications will be summarized by ATC class and preferred term. .

Anti-neoplastic medications will be coded using the WHO Drug Dictionary (WHO-DD).

Transfusions

All transfusions of blood products (incl. those not related to MDS/CMML) prior and after start of study treatment will be listed. Only MDS/CMML related transfusions (e.g. excluding transfusions due to bleeding, surgical procedure, hemolysis, infections) of platelet and red blood cells will be summarized using the FAS. For that, the number of transfusion units will be normalized by time (e.g. fixed 8-weekly interval, mentioned below as episode) prior to and on-treatment. The number of subjects with at least one transfusion episode and the number of units per episode will also be described. Further analyses to summarize transfusion independence and dependence are described in the efficacy section.

Post treatment anti-neoplastic therapies and HSCT

Anti-neoplastic medications after discontinuation of study treatment during follow-up within the study will be summarized by ATC class and preferred term. HSCTs will be also summarized with the source, the type of transplant and the allogeneic donor type.

2.5 Analysis of the primary endpoint(s)

The primary objective of the study is to compare Overall Survival (OS) between the two treatment arms.

2.5.1 Definition of primary endpoint

The primary endpoint is Overall Survival defined as time from randomization until death due to any cause, regardless of start of new therapies, HSCT, or discontinuation of treatment. If a subject is not known to have died, then OS will be censored at the latest date the subject was known to be alive (on or before the cut-off date).

The primary estimand is described by the following five attributes:

- A. The **population** is defined as higher risk MDS/CMML-2 subjects as defined by selection criteria.
- B. The primary **variable** is OS.
- C. The **treatment** is the intended study treatment of AZA+MBG/placebo, followed by any standard of care therapy including HSCT and new antineoplastic therapies.
- D. **Intercurrent events** may affect the interpretation of the variable. Based on the treatment attribute definition, OS will be analyzed regardless of treatment discontinuation, start of further anti-neoplastic therapy or HSCT.
- E. The **summary measure** is the hazard ratio (HR) and its 95% CI for OS between two treatment arms. It will be estimated using Cox proportional hazard model stratified by

randomization stratification factor. The primary comparison will be performed using log-rank test stratified by randomization stratification factor.

2.5.2 Statistical model, hypothesis, and method of analysis

Assuming proportional hazards model for OS, the following statistical hypotheses will be tested to address the primary efficacy objective for OS:

$H_{01} : \theta_1 \geq 1$ vs. $H_{a1} : \theta_1 < 1$

where θ_1 is the OS hazard ratio (MBG453+AZA versus placebo+AZA).

The analysis to test these hypotheses and compare the two treatment groups will consist of a stratified log-rank test at an overall one-sided 2.5% level of significance. The stratification will be based on the randomization stratification factor in 4 groups: i.e. Intermediate risk MDS, High risk MDS, Very high risk MDS, CMML-2.

OS analysis will be made as a part of a three-look group sequential design using a Lan-DeMets (O'Brien-Fleming) alpha spending function.

Analyses will be based on the Full analysis set according to the randomized treatment group and strata assigned at randomization. The OS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, quartiles and associated 95% confidence intervals will be presented for each treatment group. The hazard ratio for OS will be calculated, along with its 95% confidence interval, from a stratified Cox model using the same randomization stratification factor as for the log-rank test.

2.5.3 Handling of missing values/censoring/discontinuations

OS will be censored at the date of the last contact date if a subject is not known to have died at or prior to the analysis cut-off date. All deaths at or prior to cut-off date will be taken into account whenever the death occurred, i.e. even after new anti-neoplastic therapy, SCT, interruptions, or discontinuation of study treatment due to any reason.

The OS censoring reason will be summarized as:

1. Ongoing without event
2. Withdrew consent/discontinuation due to Subject or Guardian decision
3. Lost to follow-up

2.5.4 Supportive analyses

OS will be analyzed with the same statistical methods but considering the hypothetical strategy for intercurrent events of start anti-neoplastic therapy and HSCT separately:

- If death is documented after starting HSCT, the subject will be censored at the date of HSCT prior to that event rather than analyzing OS regardless of starting HSCT.
- If death is documented after starting anti-neoplastic therapy, the subject will be censored at the start of anti-neoplastic therapy rather than analyzing OS regardless of start of further anti-neoplastic therapy.

2.5.4.1 Subgroup analyses for OS

If the primary endpoint analyses for OS are statistically significant at the efficacy interim or primary analysis, subgroup analyses to assess the homogeneity of the treatment effect will be performed for the subgroups, which are specified in Section 2.2.1 . [redacted]

No formal statistical test of hypotheses will be performed. The HR together with the associated 95% CI obtained using the unstratified Cox regression model will be presented for each subgroup in a forest plot.

2.5.4.2 Sensitivity estimand analysis for primary endpoint

This sensitivity estimand will be explored only if there is a substantial proportion of patients with misclassification of IPSS-R. The two treatment arms will be compared using the calculated IPSS-R for non-CMML-2 patients. In this case, the respective calculated risk groups will be added to OS subgroup analysis (section 2.2.1).

2.6 Analysis of key secondary endpoints

The first two key secondary objectives of the study are to compare time to definitive deterioration (TTDD) of fatigue and RBC transfusion-free intervals (TFI) between the two treatment arms. Additional three key secondary objectives will be tested in a hierarchical approach described in Section 2.6.1.3.

The type I error control for this key secondary endpoint family is described in Section 2.6.1.4.

Scoring of PRO data and methods for handling of missing items will be handled according to the scoring manual and user guide for each respective subject questionnaire (Tennant 2015 , Fayers 2001 and VanReenen M and Janssen 2015).

2.6.1.1 Time to definitive deterioration (TTDD) of fatigue

Time to definitive deterioration (TTDD) of fatigue is defined as time from randomization to

1. at least 3 points worsening from baseline in FACIT-fatigue scores with no subsequently observed improvement above this threshold, or
2. death due to any cause,

whichever occurs first. A threshold of at least 3 points worsening from baseline in the FACIT fatigue score is based on Cella et al 2002.

For subjects without TTDD event, TTDD is censored at last adequate assessment. This is the last FACIT-Fatigue assessment conducted that is not missing.

For TTDD, a death event occurring after two or more consecutive missing assessments (not done or unknown) is censored in the analysis at the last adequate FACIT-Fatigue assessment before the event date and reason for censoring then summarized as ‘Death documented after two or more missing assessments’.

For those subjects who cannot deteriorate (i.e. patients with a baseline score < 3), TTDD will be censored at randomization date unless deteriorate due to death (i.e. if death occurring after

two or more consecutive missing assessments, then TTDD will be censored at the randomization date).

Assuming proportional hazards model for TTDD of fatigue, the following statistical hypotheses will be tested at the 1-sided alpha-TTDD level of significance:

H_{02} (null hypothesis): $\theta 2 \geq 1$ vs H_{a2} (alternative hypothesis): $\theta 2 < 1$

where $\theta 2$ is the TTDD hazard ratio (MBG453+azacitidine versus placebo+azacitidine).

The analysis to test this hypothesis will consist of a stratified log-rank test at the alpha-TTDD level of significance (see Section 2.6.1.4 for definitions). The same randomization stratification factor as for the OS analysis will be used.

The primary estimand is described by the following five attributes:

- A. The **population** is defined as higher risk MDS/CMML-2 subjects as defined by selection criteria.
- B. The primary **variable** is TTDD.
- C. The **treatment** is the intended study treatment of AZA+MBG/placebo, followed by any standard of care therapy including HSCT and new antineoplastic therapies.
- D. **Intercurrent events**: Based on the treatment attribute definition, TTDD will be analyzed regardless of treatment discontinuation, start of further anti-neoplastic therapy or HSCT.
- E. The **summary measure** is the hazard ratio (HR) and its 95% CI for TTDD between two treatment arms. It will be estimated using Cox proportional hazard model stratified by randomization stratification factor. The primary comparison will be performed using log-rank test stratified by randomization stratification factor.

2.6.1.2 RBC transfusion-free intervals (TFI)

RBC transfusion-free intervals (TFI) correspond to cumulative times of intervals with no evidence of RBC transfusion for at least 8 weeks at any point after randomization until death due to any cause.

Only RBC transfusions due to MDS or CMML-2 will be considered for this analysis. If the RBC transfusion-free interval is terminated by the resumption of RBC transfusion, the subjects remains "at risk", i.e. for such a subject, there can be another RBC transfusion-free interval later in time. As per the definition above, this analysis will label days where no transfusion was given as "transfusion-free" only if they fall into a transfusion-free interval lasting at least 8 weeks.

The primary outcome is the annualized RBC transfusion-free rate (TFR), which is defined as the average number of days in RBC transfusion-free intervals in a year (i.e., the total number of days in RBC transfusion-free intervals divided by the total days in the study multiplied by 365.25). The annualized RBC transfusion-free rate is estimated in a negative binomial (NB) model by using the number of days in RBC transfusion-free intervals as the response variable with natural log of time in study as an offset variable.

Under the Negative binomial (NB) model, the following statistical hypotheses will be tested at the 1-sided alpha-TF level of significance:

- H_{03} (null hypothesis): $\mu_{MBG453+azacitidine} \leq \mu_{placebo+azacitidine}$
- H_{a3} (alternative hypothesis): $\mu_{MBG453+azacitidine} > \mu_{placebo+azacitidine}$

where $\mu_{MBG453+azacitidine}$ and $\mu_{placebo+azacitidine}$ are the annualized RBC transfusion-free rates under the two treatments, respectively.

The hypotheses will be tested using a negative binomial regression model with log link, using treatment as covariate. Randomization stratification factor and baseline transfusion status (see protocol Table 8-2) will be used as stratification factors in this analysis.

The primary estimand is described by the following five attributes:

- A. The **population** is defined as higher risk MDS/CMML-2 subjects as defined by selection criteria.
- B. The primary **variable** is annualized RBC transfusion-free rate (TFR).
- C. The **treatment** is the intended study treatment of AZA+MBG/placebo, followed by any standard of care therapy including HSCT and new antineoplastic therapies.
- D. **Intercurrent events** may affect the interpretation of the variable. Based on the treatment attribute definition, annualized RBC TFR will be analyzed regardless of treatment discontinuation, start of further anti-neoplastic therapy or HSCT.
- E. The **summary measure** is the risk ratio and its 95% CI for annualized RBC TFR between two treatment arms. It will be will be tested using a negative binomial regression model with log link, using treatment as covariate. Randomization stratification factor and baseline transfusion status will be used as stratification factors in this analysis.

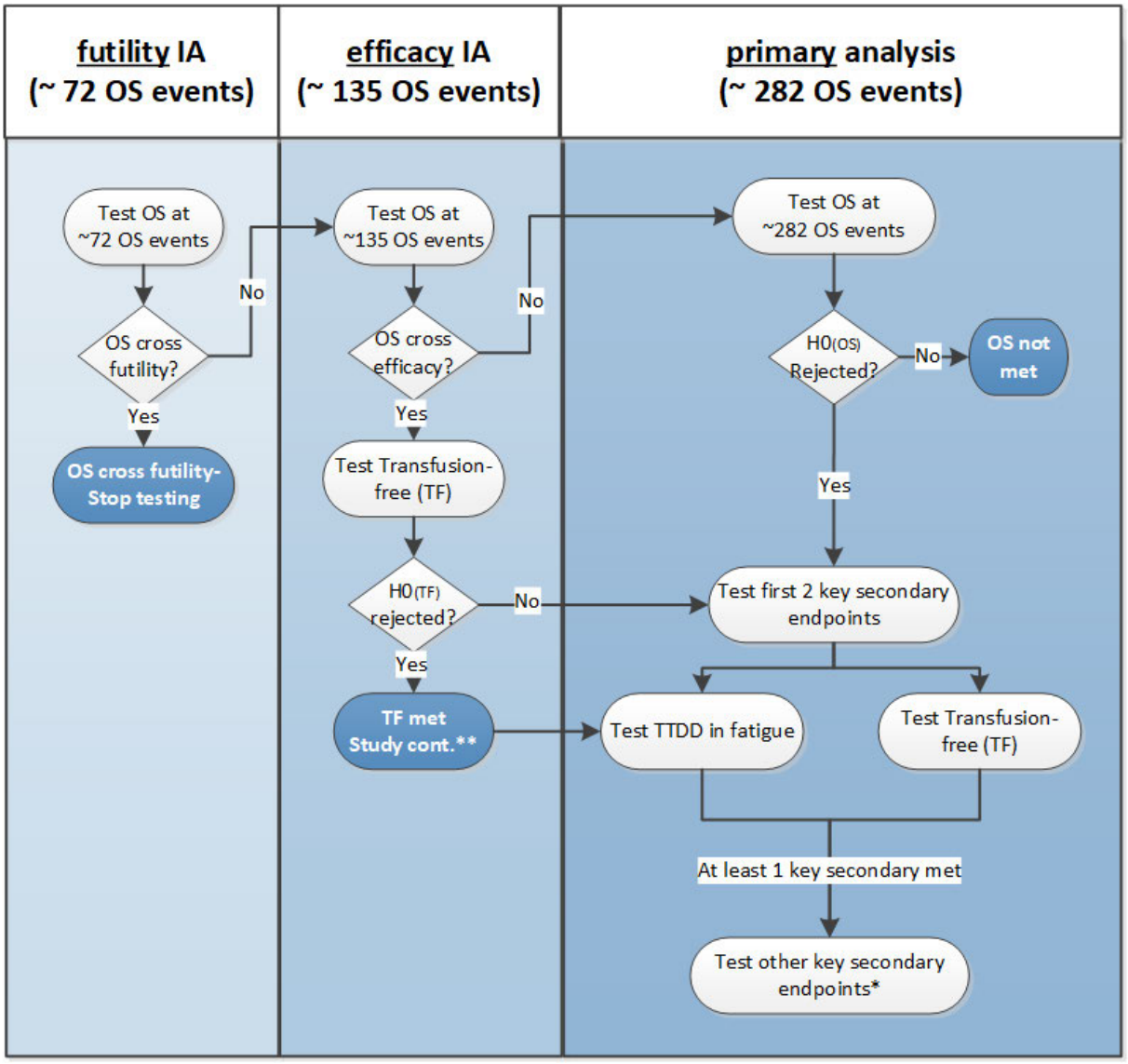
If the negative binomial regression model fails to converge, then the hypotheses will be tested using a permutation testing with the same alpha level. The parameter estimation and its 95% CI for annualized RBC TFR will be estimated using an over-dispersed Poisson model.

2.6.1.3 Hypothesis testing strategy

If OS is statistically significant at the efficacy interim or at the primary analysis, the key secondary endpoints will be tested according to the testing strategy as described in Figure 2-1:

- Time to definitive deterioration (TTDD) of fatigue
- RBC transfusion-free intervals
- Fatigue improvement rate
- Physical functioning improvement rate
- Emotional functioning improvement rate

Figure 2-1 Study flow of analyses and testing strategy



*If at least one of the two first key secondary endpoint is statistically significant, the following key secondary endpoints will be tested in a sequence, considering hierarchical testing scheme: (1) FACIT fatigue improvement rate (2) EORTC physical function improvement rate (3) EORTC emotional function improvement rate

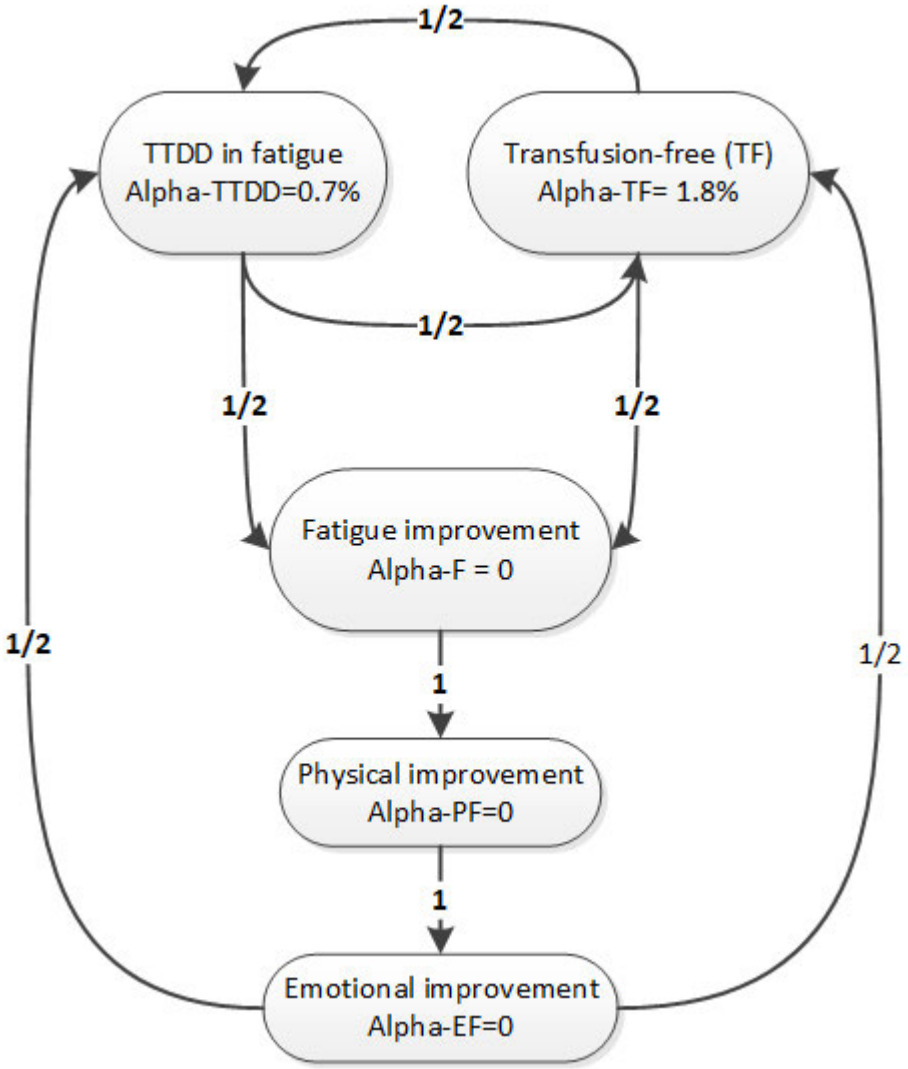
**If transfusion-free (TF) is statistically significant at the efficacy IA, then it will not be tested again at the primary analysis. The initial alpha-TF-IA assigned will be transferred to other endpoints following the gate-keeping procedure in Figure 12-2.

2.6.1.4 Gate-keeping procedure

In order to conserve the overall type I error (1-sided level of significance α = 2.5%) in this testing strategy, an alpha split with a graphical gate-keeping approach will be implemented as

shown in Figure 2-2 below based on the graphical multiple testing procedure by Bretz et al 2009, Bretz et al 2011.

Figure 2-2 Graphical gate-keeping procedure in order to control overall type I error



The alpha-TTDD of 0.7% and the alpha-TF of 1.8% are the starting values of the graphical testing procedure. The alpha-TF is the overall alpha assigned to TF, which is further split into 0.2% for the efficacy interim and 1.6% for the primary analysis.

This graphical gate-keeping procedure with the above mentioned initial alpha split and propagation ensures the protection of an overall one-sided type-I error of α = 2.5%. With exception of TF, all the other endpoints in this scheme will only be tested at the corresponding alpha's at the primary analysis.

For TF, the overall assigned alpha-TF will be split between the efficacy IA and primary analysis, with the starting value of alpha level of 0.2% at the efficacy IA and alpha level of 1.6% at the primary analysis. If TF receives alpha, it will always go entirely to the alpha level

at the primary analysis. A fixed alpha-level of 0.00001 will be spent for TTDD at the time of the efficacy interim analysis.

For example, at the efficacy interim analysis, if RBC transfusion-free intervals is statistically significant, it will not be tested again at the primary analysis. The overall alpha-TF of 1.8% will be transferred to TTDD and Fatigue improvement rate equally. Thus, at the primary analysis, TTDD will be tested at an updated starting value of alpha-TTDD 1-sided level of significance of 0.7%+0.9% = 1.6%; and fatigue improvement rate will be tested at an updated starting value of alpha-F 1-sided level of significance of 0.9 %.

2.6.1.5 PRO improvement rates

FACIT Fatigue improvement rate

For FACIT-Fatigue scale score, the responder is defined as having 3 points improvement from baseline confirmed by a second improvement of 3 points at anytime, regardless of preceding worsening. A subject who cannot improve will be considered as a non-responder.
The primary estimand is described by the following attributes:

- A. The **population** is defined as higher-risk MDS/CMML-2 subjects as defined by selection criteria.
- B. The primary **variable** is the proportion of subjects having 3 points improvement from baseline confirmed by a second improvement of 3 points at anytime.
- C. The **treatment** is the intended study treatment of azacitidine+MBG/placebo, followed by any standard of care therapy including HSCT and new antineoplastic therapies.
- D. **Intercurrent events**: Based on the treatment attribute definition, the improvement rate will be analyzed regardless of treatment discontinuation, disease progression, start of further anti-neoplastic therapy or HSCT.
- E. The **summary measure** is the proportion of responders between two treatment arms. It will be estimated using Cochran-Mantel-Haenszel method stratified by randomization stratification factor.

The following statistically hypothesis will be tested at the alpha-F level of significance:

- H_{04} (null hypothesis): $F_{MBG453+azacitidine} \leq F_{placebo+azacitidine}$
- H_{a4} (alternate hypothesis): $F_{MBG453+azacitidine} > F_{placebo+azacitidine}$

, where $F_{MBG453+azacitidine}$ and $F_{placebo+azacitidine}$ are the probabilities of fatigue responders on MBG453+azacitidine and placebo+azacitidine arms.

The proportion of responders and its 95% CI will be summarized by treatment arms. Stratified Cochran-Mantel-Haenszel method controlling for randomization stratification factor will be used to evaluate the treatment effect on the proportions of responders between treatment arms. Odds ratios with 2-sided 95% confidence interval, by randomization stratification factor and overall, will be presented. Exact method will be considered if the numbers of responders in most of the strata are low. [redacted]

EORTC - Physical functioning improvement rate

For EORTC physical functioning scale, the responder is defined as having 10 points improvement from baseline confirmed by a second improvement of 10 points at anytime, regardless of preceding worsening. A subject who cannot improve will be considered as a non-responder.

The primary estimand is described by the following attributes:

- A. The **population** is defined as higher-risk MDS/CMML-2 subjects as defined by selection criteria.
- B. The primary **variable** is the proportion of subjects having 10 points improvement from baseline confirmed by a second improvement of 10 points at anytime.
- C. The **treatment** is the intended study treatment of azacitidine+MBG/placebo, followed by any standard of care therapy including HSCT and new antineoplastic therapies.
- D. **Intercurrent events**: Based on the treatment attribute definition, the improvement rate will be analyzed regardless of treatment discontinuation, disease progression, start of further anti-neoplastic therapy or HSCT.
- E. The **summary measure** is the proportion of responders between two treatment arms. It will be estimated using Cochran-Mantel-Haenszel method stratified by randomization stratification factor.

The following statistically hypothesis will be tested at the alpha-PF level of significance:

- H_{05} (null hypothesis): $PF_{MBG453+azacitidine} \leq PF_{placebo+azacitidine}$
- H_{a5} (alternate hypothesis): $PF_{MBG453+azacitidine} > PF_{placebo+azacitidine}$

, where $PF_{MBG453+azacitidine}$ and $PF_{placebo+azacitidine}$ are the probabilities of physical functioning responders on MBG453+azacitidine and placebo+azacitidine arms.

The proportion of responders and its 95% CI will be summarized by treatment arms. Stratified Cochran-Mantel-Haenszel method controlling for randomization stratification factor will be used to evaluate the treatment effect on the proportions of responders between treatment arms. Odds ratios with 2-sided 95% confidence interval, by randomization stratification factor and overall, will be presented. Exact method will be considered if the numbers of responders in most of the strata are low.

[redacted]

EORTC - Emotional functioning improvement rate

For EORTC emotional functioning scale, the responder is defined as having 10 points improvement from baseline confirmed by a second improvement of 10 points at anytime, regardless of preceding worsening. A subject who cannot improve will be considered as a non-responder.

The primary estimand is described by the following attributes:

- A. The **population** is defined as higher-risk MDS/CMML-2 subjects as defined by selection criteria.

- B. The primary **variable** is the proportion of subjects having 10 points improvement from baseline confirmed by a second improvement of 10 points at anytime.
- C. The **treatment** is the intended study treatment of azacitidine+MBG/placebo, followed by any standard of care therapy including HSCT and new antineoplastic therapies.
- D. **Intercurrent events:** Based on the treatment attribute definition, the improvement rate will be analyzed regardless of treatment discontinuation, disease progression, start of further anti-neoplastic therapy or HSCT.
- E. The **summary measure** is the proportion of responders between two treatment arms. It will be estimated using Cochran-Mantel-Haenszel method stratified by randomization stratification factor.

The following statistically hypothesis will be tested at the alpha-EF level of significance:

- H_{06} (null hypothesis): $EF_{MBG453+azacitidine} \leq EF_{placebo+azacitidine}$
- H_{a6} (alternate hypothesis): $EF_{MBG453+azacitidine} > EF_{placebo+azacitidine}$

, where $EF_{MBG453+azacitidine}$ and $EF_{placebo+azacitidine}$ are the probabilities of emotional functioning responders on MBG453+azacitidine and placebo+azacitidine arms.

The proportion of responders and its 95% CI will be summarized by treatment arms. Stratified Cochran-Mantel-Haenszel method controlling for randomization stratification factor will be used to evaluate the treatment effect on the proportions of responders between treatment arms. Odds ratios with 2-sided 95% confidence interval, by randomization stratification factor and overall, will be presented. Exact method will be considered if the numbers of responders in most of the strata are low.

[redacted]

2.6.1.6 Supportive analysis

Supplementary estimand related to key secondary endpoints. Blinded data review identified some missing PROs/transfusions after end of treatment. Therefore 'while-on-treatment' estimand was added as supportive analysis.

TTDD of fatigue, PRO improvement rates

First supplementary estimand: the target population, the intercurrent events, and the summary measure of this endpoint are the same as for the primary estimand. The variable is based on the thresholds for improvement or deterioration as derived from anchors in this trial based on an ad-hoc analysis after database lock/unblinding (Section 2.7.2). Literature and other studies may be reconsidered when we determine the thresholds for this analysis.

Second supplementary estimand (for TTDD only): the target population, the variable, and the summary measure of this endpoint are the same as for the primary estimand. This is including all assessments with date up to and including EOT visit date. The intercurrent event of treatment discontinuation will result in censoring TTDD (i.e., to censor TTDD at the date of the last adequate assessment prior to treatment discontinuation if no TTDD event is observed prior to

the treatment discontinuation). Any on-treatment death will be considered as an event. In the summary table, this approach is referred to as '**while-on-treatment**' supplementary analysis.

Third supplementary estimand (for PRO improvement rates only): the target population, the variable, and the summary measure of this endpoint are the same as for the primary estimand. This is including all assessments with date up to and including EOT visit date. The intercurrent event will use "while-on-treatment" estimand strategy, the improvement rate will be analyzed before treatment discontinuation. In the summary table, this approach is referred to as **'while-on-treatment'** supplementary analysis.

RBC transfusion-free intervals (TFI)

First supplementary estimand: the target population, the intercurrent event, and the summary measure of this endpoint are the same as for the primary estimand. Instead of considering only RBC transfusions due to MDS or CMML-2, this variable considers all RBC transfusion records (regardless of whether due to MDS/CMML-2 or not).

Second supplementary estimand: the target population, the variable, and the summary measure of this endpoint are the same as for the primary estimand. This is including all assessments with date up to and including EOT visit date. The intercurrent event will use "while-on-treatment" estimand strategy, the RBC transfusion-free interval will only include those transfusion records before treatment discontinuation. In the summary table, this approach is referred to as **'while-on-treatment'** supplementary analysis.

2.7 Analysis of secondary endpoints

2.7.1 Efficacy

Efficacy endpoints will be calculated and summarized for the FAS.

Time-to-event endpoints will be analyzed using Kaplan-Meier method as described above. Hazard ratio and 95% CIs from the stratified Cox-model will be provided.

No formal statistical tests will be performed for any of the other secondary efficacy endpoints and hence no multiplicity adjustment will be applied.

Progression Free Survival (PFS)

PFS is defined as the time from the date of randomization to the date of the first documented progression/relapse from CR per investigator assessment or death due to any cause. Progression includes acute leukemia transformation as per WHO 2016 classification. Relapse after CR and progression will be assessed by the investigator (protocol Table 8-2). A subject without PFS event will have their PFS censored at the time of the last adequate assessment performed on or before the cut-off date.

The following events could occur after randomization and may affect the interpretation of the results:

- **Start of further anti-neoplastic therapy:** For a subject without an event before the time he/she receives any further anti-cancer therapy, PFS would be censored at the date of the last adequate assessment prior to start of further anti-neoplastic therapy.

- **Hematopoietic Stem cell transplantation (HSCT):** A subject with an event, regardless of whether it occurred before or after the time he/she receives a HSCT, would be considered as having an event at the date when the event is observed.
- **Stopping study treatment (including due to toxicities):** All events will be taken into account when they occur, regardless of any study treatment interruption or permanent discontinuation.
- **Discontinuation from study due to lost to follow-up or withdrawal of consent:** For subject without an event prior to discontinuation due to lost to follow-up or withdrawal of consent, PFS will be censored at the last adequate assessment date.

For subjects without PFS event, PFS is censored at last adequate assessment. This is the last response assessment conducted that is not considered as unknown.

For PFS, an event occurring after two or more consecutive missing response assessments (not done or unknown) is censored in the analysis of PFS at the last adequate response assessment before the event date and reason for censoring then summarized as 'Event documented after two or more missing response assessments'.

An exact rule to determine whether there are two missing assessments is therefore needed. This rule will be based on the distance between the last adequate response assessment date and the event date. If the distance is larger than threshold D_2 then the analysis will assume that there are two missing assessments. This threshold D_2 is defined as two times the protocol specified interval between the response assessments plus the allowed window around the assessments. The protocol defines that response assessment has to be done at least at the time of bone marrow assessments that are to be done every 6 cycles (24 weeks +/- 2 weeks) during the first 12 cycles and every 12 months (+/-1 month) thereafter. Therefore for the first 12 cycles, any distance larger than D_2 = 2*26 weeks = 52 weeks between last adequate assessment and the event means that there are two missing assessments. For the period beyond 12 cycles, any distance larger than D_2 = 2*13 months = 26 months (104 weeks) means that there are two missing assessments.

The PFS censoring reason will be summarized as:

1. Ongoing without event
2. New anti-neoplastic therapy
3. Withdrew consent
4. Lost to follow-up
5. Event documented after two or more missing response assessments
6. Discontinuation due to subject/physician/guardian's decision

Leukemia-free survival

Leukemia-free survival is defined as the time from date of randomization to ≥ 20% blast in bone-marrow/peripheral blood as per WHO 2016 classification or diagnosis of extramedullary acute leukemia, or death to any cause. For subject without event, the time is censored at the latest date the subject was known to be alive and without leukemia (last adequate assessment on or before the cut-off date). For time-to-event endpoints of LFS, the same censoring rules as for the PFS will be applied.

Response rate (CR/mCR/PR/HI)

Response rate is defined as the proportion of subjects with best overall response of either complete remission (CR)/marrow remission (mCR)/partial remission (PR)/hematologic improvement (HI) at any time during the study (on or before cut-off date) as per investigator assessment (for definitions, see protocol Table 82).

The best overall response (BOR) will be derived by Novartis as below:

- CR = at least one determination of CR without progression or relapse from CR within the four weeks after CR
- mCR = at least one determination of mCR (and not qualifying for a CR).
- PR = at least one determination of PR (and not qualifying for a CR/mCR).
- SD = at least one SD assessment, and not qualifying for CR, mCR or PR.
- PD = progression after randomization and not qualifying for CR, mCR, PR or SD.
- UNK = all other cases (i.e., not qualifying for confirmed CR or mCR or PR and without SD or progression).

Response rates will be provided with exact 95% confidence intervals (Clopper CJ and Pearson 1934). Subgroup analyses of response rate by blasts percentage at baseline (≤ 5% or > 5%) will be performed.

In order to be classified as CR, no progression or relapse from CR should be reported within the following 4 weeks.

In addition, the proportion of subjects with best overall response of either complete remission (CR) or partial remission (PR) at any time during the study (on or before cut-off date) as per investigator assessment will be summarized.

Stable Disease (SD) rate is defined as the proportion of subjects with best overall response of SD per investigator assessment at any time during the study (on or before cut-off date). Any hematologic improvement with exact 95% confidence intervals will be reported separately.

[redacted]

Red blood cells (RBC) / Platelet transfusion independence

RBC/Platelets transfusion independence rate is defined as the proportion of subjects having received 0 units of RBC/Platelets transfusions during at least 8 consecutive weeks after randomization. The number and percentage of subjects will be shown for the overall FAS and then also in only those with transfusion dependence at baseline (for definitions, see protocol Table 82). Percentages will be provided with exact 95% confidence intervals (Clopper CJ and Pearson 1934). Shift tables will be provided to describe the transfusion status at baseline versus the best transfusion status post-baseline.

2.7.2 Safety

For all safety analyses, the safety set will be used.

Safety summaries (tables, figures) include only data from the on-treatment period with the exception of baseline data, which will also be summarized where appropriate (e.g. laboratory shift tables). In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with start date during the on-treatment period (treatment-emergent AEs).

All safety analyses will be using the Safety Set.

On-treatment period for safety analyses

The following section streamlines the on-treatment period definition for safety analyses compared to the protocol section 12.5.2. For safety reporting, the overall observation period will be divided into three mutually exclusive segments:

1. **Pre-treatment period**: from day of subject's informed consent to the day before first administration of study treatment
2. **On-treatment period**: from day of first administration of study treatment to 30 days after last administration of study treatment (MBG453, placebo or AZA).
3. **Post-treatment period**: starting at day 31 after last administration of study treatment (MBG453, placebo or AZA).

Overall safety period: from date of first administration of study treatment to 30 days after the date of the last administration of AZA or 150 days after the last dose of MBG453 or placebo, whichever is later.

Safety follow-up period: day 31-150 after the last dose of investigational drug.

Adverse events (AEs)

AE summaries will include all AEs occurring during on-treatment period (until 30 days after last administration of study treatment). When specified, some AEs summaries will include all AEs occurring during the overall safety period.

Summary tables for adverse events (AEs) will include only AEs that started or worsened during the on-treatment period. The number (and percentage) of subjects with treatment emergent AEs will be summarized by primary system organ class, preferred term and maximum severity (based on CTCAE grades).

All AEs reported in the AE eCRF page will be listed along with the information collected on those AEs, e.g. toxicity grade, relationship to study treatment, outcome, action taken etc. AEs that started during the pre-treatment, post-treatment and after the overall safety period will be flagged.

AEs will be summarized by number and percentage of subjects having at least one AE, having at least one AE in each primary system organ class (SOC) and for each preferred term (PT). A subject with multiple occurrences of an AE will be counted only once in the respective AE category. A subject with multiple CTCAE grades for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AEs will be assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0

In the AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for

the preferred term will be based on their frequency in the MBG453 + AZA arm. The summaries will show 'All grades' (including AEs with missing grade) and 'Grades ≥ 3'.

The following adverse event summaries will be produced selecting all or a subset of AEs depending on seriousness, relationship to study treatment, outcome or action taken:

- AEs (all AEs (by SOC and by PT) and separately those considered related to study treatment
- SAEs and separately those considered related to study treatment
- SAEs with number of occurrences (an occurrence is defined as >1 day between start and prior end date of record of same preferred term).
- Non-SAEs
- SAEs with fatal outcome and separately those considered related to study treatment
- AEs leading to study treatment discontinuation
- AEs leading to dose adjustment/interruption for (MBG453/Placebo) or AZA
- AEs requiring additional therapy
- COVID-19 related adverse events by MedDRA COVID-19 (SMQ) terms.

In addition, all AEs and SAE by SOC and PT will be also provided on the overall safety period.

Adverse events for safety follow up period (day 31-150 after last dose of investigational drug) will be described by system organ class and preferred term. This analysis will include all the events newly occurred or worsened during the safety follow-up period, i.e. excluding pre-existing cases which are continuing as the same occurrence with the same or improved toxicity grade from on-treatment period but a new start date.

Occurrence: If for a same subject, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE
- more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

Separate summaries for on-treatment and all deaths (including post-treatment deaths not in the AE CRF but in the survival CRF) will be produced showing deaths reasons by SOC and preferred term. All AEs, deaths, and serious adverse events (including those from the pre- and post-treatment periods) will be listed and those collected during the pre-treatment, post-treatment and overall safety period will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened subjects.

Adverse events of special interest (AESI)

An adverse event of special interest is a grouping of adverse events that are of scientific and medical concern specific to compound MBG453. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HGLTs (high level group terms), HLT (high level terms) and PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical

concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad. These searches will be defined in the eCRS (electronic Case Retrieval Strategy) in the DMS (Document Management System) and a listing of search terms will be provided in the CSR.

For each specified AESI, number (%) of subjects with at least one event of the AESI occurring during on treatment period will be summarized together with the individual preferred terms in that grouping. In addition, number (%) of subjects with at least one AESIs by maximum CTC grade, related AESIs, serious AESIs as well as action taken and outcome of the respective AESI will be summarized.

Vital signs

Notable vital sign values during on-treatment period will be summarized using the following criteria:

Table 2-2 Notable vital sign values

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Systolic blood pressure (mmHg)	>=180 with increase from baseline of >=20	<=90 with decrease from baseline of >=20
Diastolic blood pressure (mmHg)	>=105 with increase from baseline of >=15	<=50 with decrease from baseline of >=15
Pulse rate (bpm)	>=100 with increase from baseline of >25%	<=50 with decrease from baseline of > 25%
Weight (kg)	Increase >=10% from baseline	Decrease >= 10% from baseline
Body temperature (°C)	>= 39.1	-

Clinical laboratory evaluations

Grading of laboratory values will be assigned programmatically as per National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 5.0. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account.

CTCAE Grade 0 will be assigned for all non-missing values not graded as 1 or higher.

For laboratory tests where grades are not defined by CTCAE version 5.0, results will be categorized as low/normal/high based on laboratory normal ranges.

For laboratory tests where grades are defined by CTCAE v5.0:

- Shift tables using CTCAE v5.0 grades to compare baseline to the worst on-treatment value

Liver function parameters of interest are total bilirubin, ALT, AST and alkaline phosphatase. The number (%) of subjects with worst post-baseline values as per Novartis Liver Toxicity guidelines will be summarized.

In addition for some selected hematological laboratory parameters, trends over time (baseline and on-treatment timepoints during the first 12 cycles of treatment) will be displayed via boxplots and corresponding tables displaying the summary statistics for these selected timepoints be produced, as part of the efficacy evaluations and therefore using FAS.

All CTCAE grade 3 or 4 laboratory toxicities will be listed.

ECOG Performance status

ECOG PS categorical data will be summarized by timepoint (with visit windows as defined in Table 4-1).

2.7.3 Pharmacokinetics

Pharmacokinetic (PK) data analysis will be performed for MBG453 only. The PAS will be used for all pharmacokinetic data analyses.

Descriptive statistics of PK concentration (n, m (number of non-zero concentrations), mean, coefficient of variation (CV%), standard deviation, median, geometric mean, geometric CV%, minimum and maximum) for MBG453 will be presented at each scheduled timepoint. PK parameters such as those listed in Table 2-3 will be estimated and summarized.

Below the limit of quantitation (BLQ) values will be set to zero by the Bioanalyst and will be displayed in the listing of all PK concentrations as zero and flagged. However, BLQ values will be treated as missing for the calculation of the geometric means and geometric CV% but included as zero in the other summary statistics. Missing values for any PK parameters or concentrations will not be imputed and will be treated as missing.

All concentration data for MBG453 vs. time profiles will be displayed graphically, excluding concentrations at EOT, 30-day follow-up, 150-day follow-up visit and unscheduled visit.

The concentrations collected before dose administration on Day 8 of Cycle 3 (and later cycles) are considered steady-state concentrations for MBG453.

Table 2-3 Non-compartmental pharmacokinetic parameters

Cmin* or Ctrough*	The minimum observed plasma or serum drug concentration (mass x volume-1)
Cmax* or C2h	The maximum (peak) observed plasma or serum drug concentration (mass x volume-1)

*Only steady-state PK parameters will be summarized.

Population pharmacokinetic (PopPK) analysis

If there is adequate amount of data, a mixed-effects model may be applied to the serum MBG453 concentration-time data from this study along with other studies to generate post-hoc estimates of pharmacokinetic parameters using appropriate software to characterize MBG453 exposure and to determine the effects of intrinsic (i.e. demographic factors) and extrinsic covariates (e.g. combination partners) on MBG453 exposure. If there is sufficient data for

analysis, the details of the population pharmacokinetic analyses may be provided in a separate reporting and analysis plan, and the results may be reported in a separate population pharmacokinetic report.

Immunogenicity (IG) / anti-drug antibody (ADA)

Immunogenicity will be characterized descriptively by tabulating ADA prevalence at baseline or post-baseline and ADA incidence on-treatment.

ADA incidence (i.e. ADA-positive subjects) will be calculated as the number of subjects with at least one on-treatment ADA-positive sample divided by the number of subjects with a determinant baseline IG sample and at least one determinant post-baseline IG sample.

Listings will be provided of sample ADA status (including titer for positive samples) and subject ADA status.

Details on the ADA analyses will be provided in separate document at compound level if feasible.

2.7.4 Quality of Life / Patient-reported outcomes

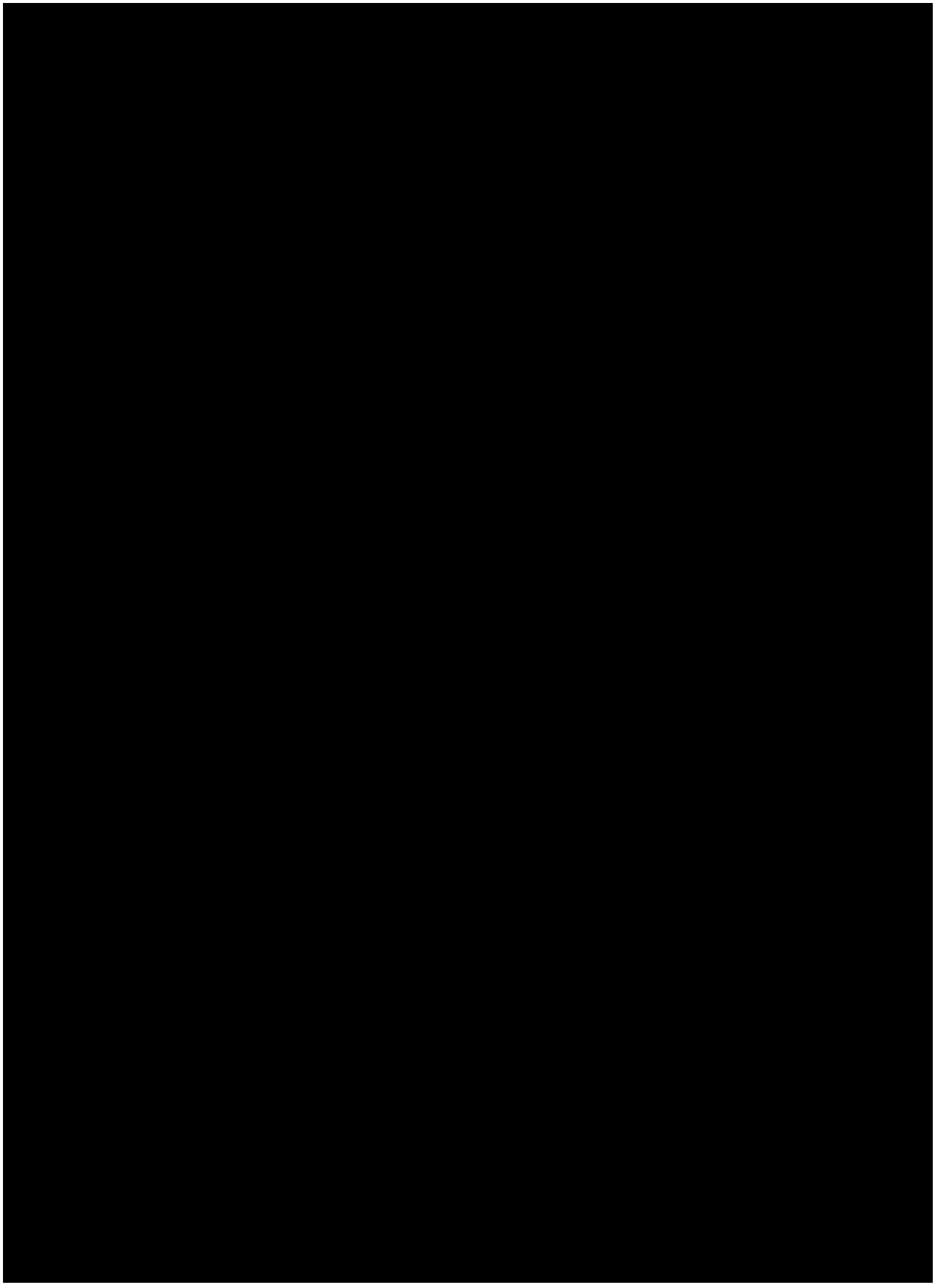
FACIT-fatigue, EORTC-physical function, EORTC-emotional function

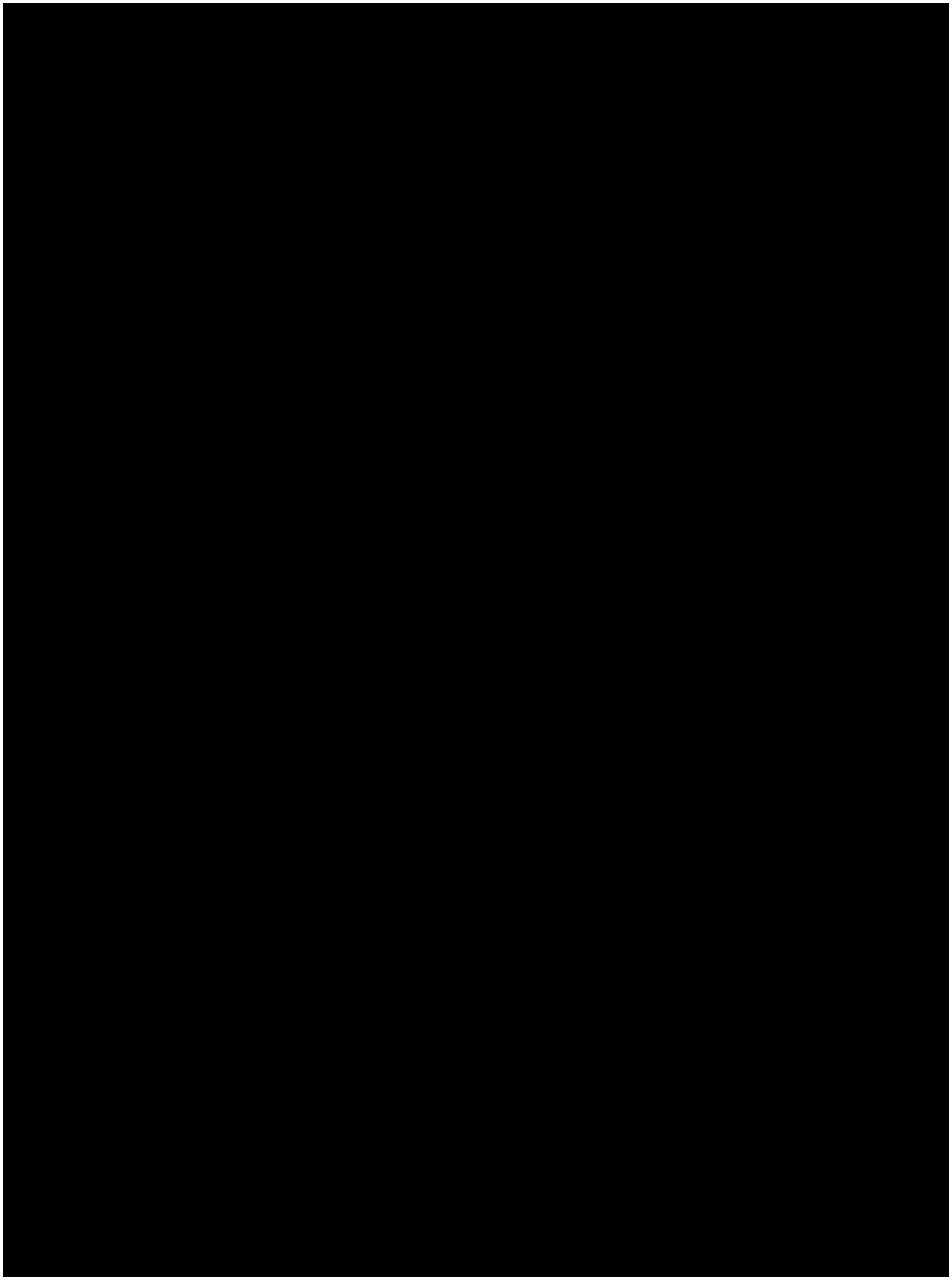
Absolute values and change from baseline in EQ-5D-5L and VAS over time will be summarized using descriptive statistics, at each scheduled assessment timepoint for each treatment arm.

Absolute values and change from baseline to Cycle 12 Day 1 in Global health status using EORTC will be summarized using descriptive statistics for each treatment arm.

Absolute values and change from baseline in FACIT-fatigue, EORTC-physical function, EORTC-emotional function over time will be summarized using descriptive statistics, at each scheduled assessment timepoint for each treatment arm.

All assessments will be taken into account whatever these assessments occurred after/before responses/progression, new antineoplastic therapy, concomitant treatment (including transfusion), SCT, interruptions or discontinuation of study treatment due to any reason.





2.9 Interim analysis

The primary efficacy analyses on overall survival will be based on the full analysis set and are event driven.

- primary analysis will be performed when approximately 282 OS events have been documented.
- futility interim analysis will be performed when approximately 40% of 180 OS events as per protocol version 00 have occurred.
- efficacy interim analysis will be performed when approximately 48% of 282 OS events have occurred and all subjects have been randomized.

Futility IA

The futility IA will be performed when there are approximately 72 OS events. Probability of Success (PoS) will be used to determine futility boundary at the futility IA. Details are specified in the statistical appendix.

At the time of the futility IA, the OS data available from the parallel ongoing Phase II study [CMBG453B12201] will be used to construct a prior to calculate the probability of success (PoS). PoS is defined as the marginal conditional probability of achieving statistical significance of the primary OS endpoint at either the efficacy interim or the primary analysis, given the Phase II and Phase III data available at the time of the Phase III futility interim. It is calculated by averaging the conditional Phase III power function over the posterior distribution for the log hazard ratio given the available Phase II and Phase III OS data. The study will be terminated for lack of efficacy after the futility IA if the PoS is less than 15%.

- If the futility IA occurs before CR rate analysis of [CMBG453B12201] study, no phase II study data will be used.

- If the futility IA occurs after CR rate analysis and before the PFS IA of [CMBG453B12201] study, OS data collected until CR rate analysis in Phase II will be used.
- If the futility IA occurs after PFS IA of [CMBG453B12201] study, OS data collected until PFS IA in Phase II will be used.

Efficacy IA

The efficacy IA will be performed when there are approximately 135 OS events and when all subjects have been randomized. Group sequential plan using Lan and DeMets α-spending approach with O'Brien-Fleming type boundary (Lan and DeMets 1983) as implemented in the software East 6.4 will be applied. The study will conclude for efficacy at the efficacy IA if the null hypothesis is rejected at one-sided significance 0.0012 (corresponding hazard ratio (HR) threshold assuming proportional hazards: HR < 0.593).

If OS is significant at the efficacy interim, it will not be tested again at the primary analysis, but the OS analysis will be updated at the primary analysis.

No external data will be used for the efficacy IA and primary analysis.

Primary Analysis

The primary analysis will be performed when there are approximately 282 OS events reported. The projected timing of interim and primary analysis is summarized in Table 2-4.

A simulation in East 6.4 of 10000 trials of the described setup of protocol version 03 (simulating for each trial the individual outcomes for the fully recruited 530 patients, using actual recruitment distribution and following group-sequential design with an interim at 135 events) was conducted and the percentage of significant trial analysis results recorded. A piecewise exponential model was used which assumed a constant hazard rate of 0.035 in both treatment arms in the first 5 months after treatment start (piece 1) and 0.021 and 0.035, respectively, (HR=0.6) subsequently (piece 2). Based on total 282 OS events at primary analysis, the cumulative probability to detect an efficacious treatment by the primary analysis is 85%, while the cumulative probability of erroneously detecting a non-efficacious treatment by the primary analysis is 2.4%.

Since the futility analysis was already conducted under protocol version 00 (40% information fraction out of the planned 180 events per protocol version 00), no changes were made to the thresholds for declaring futility or their calculation. If the primary null hypothesis as per protocol version 00 was true, then the probability to stop the trial at the futility analysis was 73%.The futility and efficacy interim analyses for OS will be performed by an independent statistician. Results from OS interim analyses will not be communicated to clinical team or any party involved in the study conduct (apart from the independent statistician and IDMC members) or external parties including Health Authority and investigators, until OS is found to be significant or study needs to be terminated due to safety or lack of efficacy.

Table 2-4 Estimated timelines for interim and primary analysis assuming 5 months delayed effects

Months after randomization of the first subject (approximation)[e]	Intent of Analysis	# OS events	Decision threshold (PoS for futility or alpha-threshold and corresponding HR for efficacy)	Boundary Crossing Probability
17	IA for Futility	72 (40% of 180 events as per protocol version 00)	15%	73%[b]
22[e]	IA for Efficacy	135 (48% of 282 events as per protocol version 03)[a]	α<0.0012 (HR <0.593) [d]	33%[c]
44[e]	Primary	282 (100% of 282 events as per protocol version 03)[a]	α<0.0246 (HR <0.791) [d]	85%[c]

[a] Efficacy interim and primary analysis using O'Brien-Fleming boundaries in Group Sequential Design
[b] The futility boundary crossing probability was calculated under null hypothesis in Phase III and under null hypothesis in Phase II, based on the assumption that OS data from CR analysis of [CMBG453B12201] was available at the time of futility interim analysis.
[c] The efficacy boundary crossing probability is calculated under alternative hypothesis (HR=1 for the first 5 months followed by HR=0.6) using East 6.4
[d] Alpha-thresholds and corresponding HR thresholds (assuming proportional hazards) are calculated using East 6.4
[e] timeline estimated based on actual recruitment pattern

2.10 Sample size calculation

2.10.1 Primary endpoint(s)

The sample size calculation is based on the primary variable OS. The hypotheses to be tested and details of the testing strategy are described in protocol Section 12.4.2.

With 282 OS events and assuming a delayed onset of effect of 5 months followed by HR=0.6, the cumulative probability to detect an efficacious treatment by the primary analysis is 85% based on 10000 simulations in East 6.4 using the piecewise exponential model described in Section 12.7. The median OS for Placebo+HMA is assumed to be 20 months and actual recruitment pattern at time of protocol version 03 was used.

This calculation assumes analysis by a one-sided log-rank test at the overall 2.5% level of significance where subjects are randomized to the two treatments in a 1:1 ratio. Assuming that total enrollment time will be 20 months (based on the observed recruitment rate), and losses to follow-up of 10% per treatment arm per year, a total of 530 subjects will need to be randomized to observe the targeted 282 OS events at about 44 months after the randomization date of the first subject. These calculations were made using the software package East 6.4 and R 3.4.3.

The efficacy analyses on overall survival will be based on full analysis set and are event driven. Primary analysis will be performed when there are approximately 282 OS events. A futility interim analysis will be performed when approximately 40% of the 180 OS events as per protocol version 00 have occurred. A efficacy interim analysis will be performed when approximately 48% of the 282 OS events have occurred and all subjects have been randomized.

2.10.2 Secondary endpoint(s)

The hypotheses to be tested and details of the testing strategy for the two key secondary endpoints are provided in Section 2.6. No power consideration are made for other key secondary endpoints.

Time to definitive deterioration (TTDD) of fatigue

Based on available data from previous study [CETB115D2301], out of the subjects remaining on study, the median of FACIT fatigue showed a worsening of 3 point from baseline by Cycle 14. Taking into account that the Kaplan-Meier first quartile estimate of OS in the azacitidine+placebo was 273 days (about 9 months), we assume the median TTDD of fatigue as 9 months for azacitidine+placebo arm in this study. It is hypothesized that treatment with MBG453+ azacitidine will result in a 40% reduction in the hazard ratio for TTDD, i.e. an expected hazard ratio of 0.6 which corresponds to a median TTDD of fatigue of 15 months in MBG453+ azacitidine arm under the exponential model assumption.

If OS is significant at the primary analysis, and assuming an HR of 0.6 for TTDD, it is calculated that approximately 180 TTDD events are observed which ensures 90% power at the primary analysis with an alpha-TTDD level of 1.6%.

RBC Transfusion-free intervals

Based on available data from previous study [CETB115D2301], out of 164 subjects randomized in azacitidine+placebo arm, from Cycle 1 to Cycle 4, 123 (75%) subjects needed RBC transfusions, with a median of 5 RBC transfusions and standard deviation of 5.06.

We assume the over-dispersion parameter (k=0.5) and the annualized RBC transfusion-free rate for subjects treated with azacitidine+placebo is $\mu_{placebo+azacitidine} = 0.25$, the annualized RBC transfusion-free rate for those treated with MBG453+ azacitidine is $\mu_{MBG453+azacitidine} = 0.385$. Therefore, the estimated annualized transfusion-free rate increase for MBG453+azacitidine versus azacitidine+placebo is 35%. A total number of 500 subjects will give overall 91% statistical power at a 1-sided significance level of 0.016 given OS is significant.

3 Change to protocol specified analyses

No shift table using the low/normal/high/ (low and high) classification to compare baseline to the worst on-treatment value will be done. However, abnormal values are flagged in listings and flag for normal range comparison is stored in the datasets.

4 Appendix – General definitions

Date of first administration of investigational drug/AZA

The date of first administration of investigational drug/AZA is defined as the first date when a non-zero dose of investigational drug/AZA is administered and recorded on the study treatment eCRF page.

Date of last administration of investigational drug/AZA

The date of last administration of the investigational drug/AZA is defined as the last date when a non-zero dose of the respective investigational drug/AZA is administered and recorded on the study treatment eCRF page. So both, first and last date of investigational drug/AZA, are derived separately for each drug which is part of the study treatment.

Date of first administration of study treatment (combination)

The date of first administration of study treatment (or start date of study treatment) is defined as the first date when a non-zero dose of any component of the study treatment (MBG453/Placebo or azacitidine) is administered.

Date of last administration of study treatment (combination)

The date of last administration of study treatment (or end date of study treatment) is defined as the last date when the last non-zero dose of any last component of the study treatment (MBG453/Placebo or azacitidine) is administered.

Date of last exposure to AZA and investigational drug and date last exposure to study treatment (combination)

One planned cycle length is 28 days. Azacitidine are planned to be administered every cycle on days 1 to 7 (or on days 1 to 5 and days 8 and 9). MGB453/placebo are administered every 4 weeks (Q4W), on Day 8 of each cycle, unless there was a toxicity leading to a dosing interval increase.

As treatment is given in cycles, the date of last administration of study treatment is not considered the date of last exposure but instead the planned end date of the last cycle in which the last non-zero dose was given.

The **date of last exposure to MBG453/placebo** is therefore calculated as:

- Minimum (last date of administration of investigational drug + 27 days, date of death, last contact date in case subject is lost to follow-up, cutoff date), as MBG453/placebo injections are given Q4W (in case of change to Q8W, the cycle would end after 55 days)

The **date of last exposure to AZA** is calculated as:

- Minimum (last date of administration of AZA in the last cycle + 20 days, date of death, last contact date in case subject is lost to follow-up, cutoff date) as AZAs are given in the first week of each cycle followed by a period of at least 3 weeks prior to the next cycle (total cycle length 28 days)

The **date of last exposure to study treatment (combination)** is calculated as the minimum date between the latest of last date of exposure to any component of the study treatment, the date of death, the last contact date in case subject is lost to follow-up, and the cutoff date. The end date of the last cycle initiated is the planned end date (Day 28) of the last cycle initiated where MBG453 and/or AZA were last administered or the date of the last administration of a non -zero dose of any last component of the study treatment (MBG453, AZA), whichever is the

latest. The duration of exposure to study treatment (combination) will be calculated as follows: date of last exposure to study treatment (combination) - date of first administration of study treatment (combination) + 1.

The planned end date of a cycle (Day 28) will be applicable even if this date goes beyond the data cutoff date (it should not be truncated to the date of data cutoff).

Duration of study treatment (combination) used to calculate the dose intensity and the relative dose intensity

For the calculation of the dose intensity and relative dose intensity, the cycle-based exposure to study treatment (combination) will be derived from the date of first administration of study treatment (combination) to the actual or planned end date of the last cycle initiated (whichever is the latest) irrespective of date of death, last contact date for withdraw consent and cut-off date.

- The **start date of the first cycle initiated** is the first date when a non-zero dose of any component of the study treatment (MBG453/Placebo or azacitidine) is administered.
- The **end date of the last cycle initiated** is the planned end date (Day 28) of the last cycle initiated where MBG453/Placebo and/or azacitidine was administered, or the date of last administration of a non-zero dose of any component of the study treatment (MBG453/Placebo or azacitidine), whichever is the latest.

The duration in days from first to last cycle initiated will be calculated as follow: end date of the last cycle initiated - start date of the first cycle initiated + 1.

The duration of study treatment in days used to derive the dose intensity and the relative dose intensity will be calculated as follow: end date of the last cycle initiated – date of first administration of study treatment (combination) + 1.

The planned end date of a cycle (Day 28) will be applicable even if this date goes beyond the data cutoff date, the date of death or the date of last contact date for withdraw consent (it should not be truncated to the date of data cutoff, the date of death or the date of last contact date).

Study day

The study day, describes the day of the event or assessment date, relative to the reference start date (randomization date or start date of study treatment).

The study day is calculated as follows:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

The reference start date for safety assessments (e.g. adverse event onset, laboratory assessment, vital sign measurement etc.) is the start of study treatment. The reference start date for efficacy (e.g. response assessment, time-to-event endpoints) and patient-reported outcomes (PRO) is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Time unit

A year length is defined as 365.25 days. A month length is 30.4375 days (365.25/12). If duration is reported in months, duration in days will be divided by 30.4375. If duration is reported in years, duration in days will be divided by 365.25.

Baseline

For efficacy evaluations and PROs, the last non-missing assessment, including unscheduled assessments on or before the randomization date is taken as "baseline" value or "baseline" assessment.

For safety evaluations, the last available assessment including unscheduled assessments on or before the date of start of study treatment is taken as ''baseline" assessment.

If subjects have no value as defined above, the baseline result will be missing.

Windows for multiple assessments

Time windows will be defined for descriptive summary of ECOG and PRO data (EORTC QLQ-C30, EQ-5D-5L, FACIT-Fatigue, [redacted]) by visit and the longitudinal data analysis. If more than one assessment is available in the same time window, the assessment closest to the planned date will be considered. If two assessments are obtained with the same time difference compared to the scheduled visit day, the assessment obtained prior to the visit will be considered. Data obtained at the end of treatment visit will be classified as other assessment in the corresponding time window. For subjects who discontinue treatment for reasons other than progressive disease, withdraw consent or death, EORTC QLQ-C30, FACIT-Fatigue and EQ-5D-5L will continue being collected every 12 weeks during this follow-up period until death.

Table 4-1 Time windows for PRO and ECOG data

Assessment	Target day of assessment	Time Window Definition
Baseline	On or before Study Day 1[a]	≤ Study Day 1
During treatment phase (an example for PRO assessments done every cycle, need to be adjusted for each questionnaire based on the assessment schedule)		
Cycle 2 Day 1	Study Day 29	Study Days 2 to 42
Cycle 3 Day 1	Study Day 57	Study Days 43 to 70
Cycle k day 1 (with k=4, 5, 6, 7, 8 etc...)	Study Day = 28*(k-1) + 1	Study Day 28*(k-1) +1 -14 to 28*(k-1) +14 For last cycle of dosing : from 28*(k-1) +1 -14 to end of treatment visit date +7 "Note: EOT data will be included if obtained within 7 days of permanent discontinuation of study treatment"

Assessment	Target day of assessment	Time Window Definition
After treatment discontinuation (only EORTC QLQ-C30, FACIT-Fatigue and EQ-5D-5L)		
Post treatment follow-up	Every 12 weeks	For the first time window: [upper bound of the last previous time windows with assessment + 1; PRO assessment date + 42 days] Otherwise: [PRO assessment date /+ 84 days]
Survival follow-up	Every 12 weeks	For the first time window: [upper bound of the last previous time windows with assessment + 1; PRO assessment date + 42 days] Otherwise: [PRO assessment date /+ 84 days]
[a] Study Day 1 = randomization date will be baseline ([redacted]). If C1D1 assessment was performed after randomization it is still considered as 'baseline' if performed within 7 days of randomization and on or before the start date of any study treatment.		

Last contact date

The last contact date will be used for censoring of subjects in the analysis of overall survival.

The last contact date is defined as the latest complete date from the below list on or before the data cut-off date (Table 4-2). The cut-off date will not be used for last contact date, unless the subject was seen or contacted on that date. No date post cut-off date will be used.

Completely imputed dates (e.g. the analysis cut-off date programmatically imputed to replace the missing end date of a dose record) will not be used to derive the last contact date. Partial date imputation is allowed for event (death)/censoring only if coming from the 'Survival' eCRF.

The last contact date will be derived for subjects not known to have died at the analysis cut-off using the last complete date among the following:

Table 4-2 Last contact date data sources

Source data	Conditions
Date of Randomization	No condition
Last date subject was known to be alive from Survival Follow-up page	Subject status is reported to be alive or unknown
Start/End dates from further antineoplastic therapy	Non-missing medication/procedure term
Start/End dates from drug administration record	Non-missing dose
Response assessment date	Response marked as 'done'
Laboratory/PK collection dates	Sample collection marked as 'done'
Vital signs date	At least one non-missing parameter value
ECOG performance status date	Non-missing performance status
Start/End dates of AE	Non-missing verbatim term

IPSS-M derivation

IPSS-M classification are derived based on those MDS patients with baseline mutation data available.

Baseline mutation status was evaluated using a targeted 37 gene NGS panel on genomic DNA extracted from BMMC and/or PB samples, with a sensitivity of 2% VAF.

IPSS-M was derived based on Bernard et al, 2022 (https://github.com/papaemmelab/ipssm) using available clinical data (BM blast, HBG, Platelet), cytogenetic categories, and baseline mutation. Data on the following genetic alterations was missing del17/17p, TP53 loss-of-heterozygosity (LOH), BCORL1, GNB1, NF1, PPM1D, PRPF8 and MLL-PTD. Considering low prevalence of MLL-PTD, this alteration was considered as non-mutated for all patients to minimize the number of patients with undetermined IPSS-M score. All other missing variables were considered as missing (i.e., imputed as NA in the dataset).

5 References

Bretz F, Maurer W, Brannath W, et al (2009) A graphical approach to sequentially rejective multiple test procedures. Stat Med; 28(4): 586-604.

Bretz F, Posch M, Glimm E, et al (2011) Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biom J; 53(6): 894-913.

Cheson BD, Bennett JM, Kantarjian H, et al (2000) Report of an international working group to standardize response criteria for myelodysplastic syndromes. Blood; 96(12): 3671-4.

Cheson BD, Greenberg PL, Bennett JM, et al (2006) Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. Blood; 108(2): 419-25.

Clopper CJ and Pearson ES (1934) The use of confidence or fiducial limites illustrated in the case of the binomial. Biometrika; 26(4): 404-413

Fayers PM, Aaronson NK, Bjordal K, et al (2001). On behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by European Organisation for Research and Treatment of Cancer, Brussels.

Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al (2009) Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol; 10(3): 223-32.

Kim K, DeMets DL (1987) Design and analysis of group sequential tests based on the type I error spending rate function. Biometrika; 74(1): 149-154.

Platzbecker U, Fenaux P, Adès L, et al (2018) Proposals for revised IWG 2018 hematological response criteria in patients with MDS included in clinical trials. Blood; [Epub ahead of print].

Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials. Biometrika; 64(2): 191-99.

Tennant, K. (2015). Assessment of fatigue in older adults: the FACIT Fatigue Scale (version 4). Supportive Care in Cancer, 23(5), 1355-1364.

VanReenen M and Janssen B (2015) EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument.

Silverman LR, Demakos EP, Peterson BL, et al (2002) Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B J Clin Oncol; 20(10): 2429-40.