16.1.9 Statistical Analysis Plan

Statistical Analysis Plan (31 August 2020)

SGI-110-07

Statistical Analysis Plan

A Phase 3, Multicenter, Randomized, Open-Label Study of Guadecitabine (SGI-110) versus Treatment Choice in Adults with Myelodysplastic Syndromes (MDS) or Chronic Myelomonocytic Leukemia (CMML) Previously Treated with Hypomethylating Agents

Date: 31 August 2020

Based on: Protocol Version 5.0

Authored by:

[redacted] Study Biostatistician

[redacted]
1514E5155DA54447A5B6F02C256276C5

Approved by:

[redacted]

[redacted]

[redacted] Clinical Development

[redacted]

[redacted] Clinical Development

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND DEFINITIONS

ADaM	analysis data model
AE	adverse event
AML	acute myelogenous leukemia
ANC	absolute neutrophil count
BDM	Biostatistics and Data Management
BID	twice daily
BM	bone marrow
BSA	body surface area
BSC	best supportive care
CI	confidence interval
CMML	chronic myelomonocytic leukemia
C_{max}	maximum concentration
CR	complete response
CRO	contract research organization
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EQ VAS	EQ visual analogue scale
FAB	French-American-British
HCT	hematopoietic cell transplant
Hgb	hemoglobin
HR	hazard ratio
IC	standard intensive chemotherapy
ITT	intent-to-treat
IWG	International Working Group
LDAC	low-dose Ara-C (cytarabine)
MDS	myelodysplastic syndromes
MedDRA	Medical Dictionary for Regulatory Activities
NCCN	National Comprehensive Cancer Network
NDAOH	number of days alive and out of the hospital
NE	nonevaluable
NR	nonresponders
OR	odds ratio
OS	overall survival
PB	peripheral blood
PD	pharmacodynamic(s)
[redacted]	[redacted]
PR	partial response
PT	preferred term
QOL	quality of life
QTc	heart rate corrected QT interval
ROW	rest of world
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
SGI-110	guadecitabine
SOC	system organ class
TC	treatment choice
WBC	white blood cell

1.0 INTRODUCTION

This Statistical Analysis Plan (SAP) is based on SGI-110-07 protocol version 5.0, dated 02 October 2018. Analyses and statistical reporting for SGI-110-07 will be conducted by Astex Pharmaceuticals Biostatistics department with the interim analysis results as reported by the study's Data Monitoring Committee (DMC). The analyses specified in this document supersede the high-level analysis plan described in the protocol.

1.1 Background of the Disease

Myelodysplastic syndromes (MDS) are a heterogeneous group of hematopoietic stem cell disorders characterized by dysplastic changes in myeloid, erythroid, and megakaryocytic progenitors. This group of diseases clinically presents with and is associated with cytopenias affecting one or more of the three lineages and may progress to acute myeloid leukemia (AML) (Bennett et al 1982; Cheson et al 2000; Heaney and Golde 1999; Silverman 2003). While MDS is relatively uncommon in the general population, the incidence increases with age, hence disease management is often complicated by presence of non-malignant co-morbidities and lessened ability to tolerate intensive treatment.

Chronic myelomonocytic leukemia (CMML) is a clonal hematologic malignancy characterized by an absolute peripheral monocytosis, ineffective hemopoiesis, splenomegaly, and an increased risk of transformation to AML. The median age at diagnosis is 65 to 75 years, with disease predominance in males. Prognosis is generally poor, with a median survival of 12 to 19 months (Padron et al 2014). Genetic studies have also shown that CMML has a distinct pattern of associated genetic abnormalities (Meggendorfer et al 2012).

Patients die either from complications associated with cytopenias (infections and bleeding) or from transformation to AML.

1.2 Treatment Options

All patients with MDS and CMML receive supportive care (eg, growth factors, blood or platelet transfusions, antibiotics). High-dose chemotherapy with stem cell/bone marrow transplant is the only potentially curative treatment. However, most patients (older patients or patients who have other medical problems) are not candidates for a higher-risk treatment such as stem cell transplant. The intensity of treatment (chemotherapy including hypomethylating agents [HMAs]) and ability to undergo stem cell transplant is largely based on risk factors. A combination of these therapies is often used. However, with the discovery and use of treatments such as azacitidine, decitabine, and lenalidomide, the treatment paradigm for MDS has been transformed (Bejar and Steensma 2014).

Two hypomethylating agents (HMAs) have been approved in most countries for the treatment of intermediate and high risk MDS and CMML: azacitidine and decitabine. These agents have dramatically changed both the course of treatment for MDS and have improved the outcome of patients who previously had very poor survival.

Treatment for CMML is similar to MDS. As with MDS, the sole curative therapy for CMML currently has been shown to be allo-HSCT. However, most CMML patients are not candidates for transplant because of advanced age and/or co-morbidities. The HMAs azacitidine and decitabine are also currently approved for the treatment of CMML in several jurisdictions.

Standard front-line treatment for MDS patients with intermediate or high risk is HMAs (azacitidine or decitabine). Patients who are candidates for stem cell transplant and have a donor available undergo transplant. Patients who do not respond to treatment with HMAs have few options and no options that have been shown to be effective or to prolong survival after failure of HMAs.

At present, there are no approved treatment options for MDS patients who rapidly progress on HMA therapy or who fail to respond to adequate HMA treatment. Other than hypomethylating agents, common "conventional care" regimens in MDS patients include low-dose cytarabine (LDAC), standard intensive chemotherapy (IC) of a 7+3 regimen, or best supportive care (BSC) (Fenaux et al 2009; Seymour et al 2010).

1.3 Guadecitabine

Guadecitabine is a new chemical entity that incorporates decitabine and deoxyguanosine linked by a phosphodiester bond. Unlike decitabine, guadecitabine is resistant to deamination by cytidine deaminases (CDAs). Compared with IV decitabine, decitabine from subcutaneous (SC) guadecitabine has prolonged exposure and lower C_{max} (Issa et al 2015). This differentiated pharmacokinetic (PK) profile is the proposed basis for potential enhancement of clinical activity of guadecitabine.

2.0 STUDY OBJECTIVES

2.1 Primary Objective

To assess and compare overall survival (OS) between guadecitabine and treatment choice (TC) in adults with MDS or CMML previously treated with a hypomethylating agent (azacitidine or decitabine, or both).

2.2 Secondary Objectives

To assess and compare effects of guadecitabine and TC in adults with MDS or CMML previously treated with a hypomethylating agent (azacitidine or decitabine, or both) with respect to the following variables:

- Transfusion independence.
- Marrow CR (mCR) with transfusion independence.
- Survival rate at 1 year after randomization.
- Leukemia-free survival.
- Number of days alive and out of the hospital (NDAOH).

- Disease response based on IWG 2006 MDS criteria, including CR, partial response (PR), marrow CR (mCR), and hematological improvement (HI; erythroid, platelet, or neutrophil response).
- Duration of response.
- Number of red blood cell (RBC) and platelet transfusions.
- Health-related quality of life (QOL).
- Safety.

[redacted]

3.0 STUDY DESIGN

3.1 Overall Study Design

This is a Phase 3, randomized, open-label, parallel-group, multicenter study designed to evaluate the efficacy and safety of guadecitabine in subjects with MDS or CMML who failed or relapsed after adequate prior treatment with azacitidine, decitabine, or both. Approximately 408 subjects from approximately 100 study centers will be randomly assigned in a 2:1 ratio to either guadecitabine (~272 subjects) or TC (~136 subjects):

- Guadecitabine: 60 mg/m^2 given SC daily on Days 1-5 in 28-day cycles (delayed as needed to allow blood count recovery).
- TC (low-dose cytarabine, standard intensive chemotherapy and best supportive care).

Before randomization the investigator will assign each subject to one of the following TC options based on the subject's prior treatment received, country approval, and local institutional standards:

- Low-dose cytarabine (LDAC).
- Standard intensive chemotherapy (IC) of a 7+3 regimen.
- Best supportive care (BSC) only.

Randomization will be stratified by disease category (MDS vs CMML), BM blasts (BM blasts >10% vs BM blasts ≤10%), preselected TC option (LDAC vs IC vs BSC), and study center region (North America vs the rest of the world [ROW]).

Subjects with MDS and CMML may expect a different treatment outcome, with CMML subjects experiencing a better treatment response and longer survival. Subjects who have higher baseline percentage of BM blasts tend to have a worse survival than subjects with low BM blast percentage

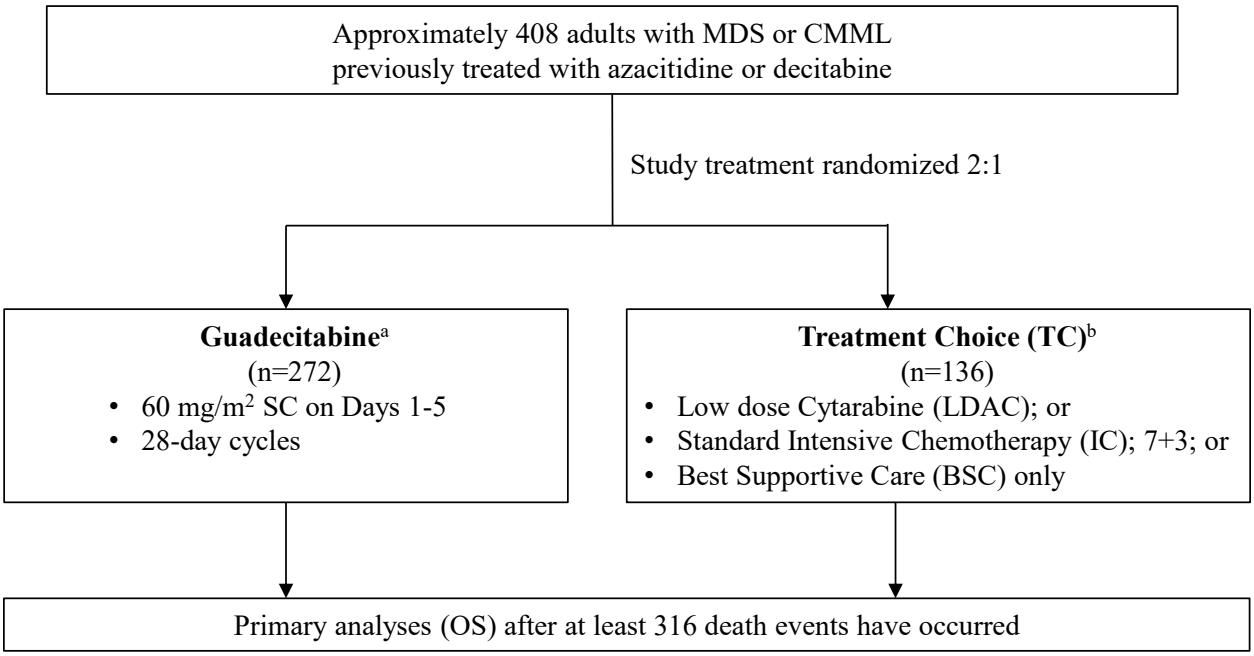
at baseline. Similarly, subjects with a different selection of TC may expect a different treatment outcome. In addition, prognosis could have regional differences due to multiple factors including variable levels of standard of care. Stratification by these known and important prognostic factors measured at baseline prevents imbalance between treatment groups within strata, and could reduce bias and improve power of the study.

Peripheral blood (PB) will be collected at screening and on Days 1, 8, 15, and 22 of Cycles 1 and 2 and on Days 1 and 15 for Cycles 3 to 6. Bone marrow (BM) aspirate or biopsy will be performed at screening and at the end of Cycles 2, 4, and 6. Subjects treated with standard IC (TC arm) may have an additional BM aspirate or biopsy after the first cycle of induction in addition or instead of the end of Cycle 2 BM assessment. In cycles where BM is not required, the Day 1 PB and the most recent prior BM data will be used for the response assessment. After Cycle 6, BM assessment (BM aspirate or biopsy) will be repeated every 4 months until PB or BM assessment shows disease progression or relapse.

The sponsor, investigators, and study subjects are not blinded in this study.

Figure 1 below summarizes the study design.

Figure 1: Study Schema



[a] Treatment with guadecitabine should continue for at least 6 total cycles in the absence of unacceptable toxicity or disease progression requiring alternative therapy. Beyond 6 cycles, treatment should continue as long as the subject continues to benefit.

[b] Before randomization the Investigator will assign each subject to one of the TC options.

3.2 Study Endpoints

3.2.1 Primary Endpoint

- OS, defined as the number of days from the day the subject was randomized to the date of death (regardless of cause).

3.2.2 Secondary Endpoints

- Transfusion independence for any 8 consecutive weeks based on rolling 8-week assessments. Number and rate of subjects who are free of transfusions for 8 consecutive weeks (or more) at any time after start of study treatment and who have hemoglobin (Hgb) of ≥8 g/dL and platelets ≥20,000/μL.
- Marrow CR (mCR) based on IWG 2006 criteria with transfusion independence (as defined above).
- Survival rate at 1 and 2 years after randomization.
- Leukemia-free survival defined as the number of days from randomization to the date when BM or PB blasts reach ≥20%, or death of any cause.
- Number of days alive and out of the hospital (NDAOH).
- Disease Response including CR, mCR, PR, and HI; including HI with erythroid [HI-E], neutrophil [HI-N], or platelet [HI-P] response, based on IWG 2006 criteria.
- Duration of response.
- Number of RBC or platelet transfusions (units) over the duration of the study treatment.
- Health-related QOL by EQ-5D (consisting of the EQ-5D-5L descriptive system and the EQ visual analogue scale [EQ VAS]).
- Incidence and severity of AEs.
- 30- and 60-day all-cause mortality.

4.0 SAMPLE SIZE

In order to provide power of at least 89% to detect a difference in hazard ratio of approximately 0.68 (median OS of 6 months for the TC control arm and 8.8 months for the guadecitabine arm) using a 2-sided stratified log-rank test at an overall two-sided 0.05 alpha level, given the use of a 2:1 randomization, the trial will require 316 death events. Assuming accrual is uniform over an 18-month enrollment period with an additional follow-up of 9 months, approximately 408 patients will need to be randomized.

If, after a follow-up of 12 months from the last subject randomized, the 316 death events have not occurred, the primary analysis will be conducted at 12 months from the last subject randomized if 277 or more deaths have occurred. If at 12 months <277 deaths have been observed the primary analysis will be conducted as soon as 277 death events have been observed (corresponding to 85% power).

5.0 ANALYSIS SETS

5.1 Efficacy Analysis Set

Efficacy analyses will be based on the intent-to-treat (ITT) principle. The Efficacy Analysis Set will include data from all subjects randomly assigned to study treatment. All data will be included and no subjects excluded because of protocol violations. For the analysis of efficacy data, subjects will be included in the treatment group according to their randomly assigned treatment.

For NDAOH and health-related QOL by EQ-5D-5L, the primary analysis will include only the data collected during the first 6 months of the study because, during this study period, subjects are assessed monthly whether or not they are still on study treatment.

5.2 Safety Analysis Set

The Safety Analysis Set will include data from all subjects randomly assigned to study treatment who receive any amount of study treatment or any component of a multi-dose study treatment regimen. All data will be included and no subjects excluded because of protocol violations.

For safety data analysis, subjects will be included in the treatment group according to the treatment they actually receive. If a subject receives more than one type of study treatment during the study, the subject will be included the treatment group according to the treatment received most frequently. In the unlikely event that a subject receives the same number of cycles of two different treatments the subject will be included in the treatment group according to the treatment received first.

5.3 PK Analysis Set

[redacted]

5.4 Biomarker Analysis Set

[redacted]

6.0 SCHEDULE OF ANALYSES

Data listings and summary tables are reviewed by the DMC approximately every 6 months to ensure the safety of study subjects and to enhance the quality of trial conduct (refer to protocol Section 4.4 and the DMC Charter). These data listings and summary tables are generated by Axio Research LLC, an independent contract research organization (CRO) supporting the DMC activities.

One formal interim analysis of OS was planned in the protocol after approximately half (~158) of the required 316 death events had occurred. This interim analysis was conducted by an independent DMC on 16 April 2019, resulting in a recommendation to the Sponsor that the study should continue.

At Sponsor's request, an additional earlier interim analysis of futility was conducted by the DMC at one of the scheduled DMC meetings on 12 September 2018 when ~100 death events had occurred. This interim analysis was added due to emerging information on failure in efficacy of a similar trial conducted by the Sponsor in treatment-native AML subjects using the same drug product. This meeting resulted in a recommendation to the Sponsor that the study should continue.

To achieve a mature analysis of OS and also keep the study duration within a reasonable period, the primary analyses of all study data will be performed either after 316 death events (corresponding to a 89% power) have occurred within 12 months from the date of the last subject randomization, or after 277 death events (corresponding to an 85% power) have been observed if 316 death events were not reported yet by 12 months from the date of the last subject randomization.

7.0 STATISTICAL ANALYSIS

Unless otherwise specified, treatment comparisons will be performed between the guadecitabine group and the TC group. [redacted]. Procedures will be employed to control the alpha error associated with testing the multiple endpoints (see Section 7.3.3). The statistical test to be employed for each comparison is specified below. The SAS® statistical package (SAS Institute Inc., Cary, NC, USA, version 9.4 or a later version) will be used for the analyses.

The following data listings by study center and subject will be provided, as recommended by the ICH E3 guideline "Structure and Content of Clinical Study Reports": discontinued subjects, protocol deviations, demographics, individual efficacy data, subjects excluded from the efficacy analysis, AEs, medications, and relevant laboratory measurements. Additional data listings may be generated to support other relevant discussions in CSR.

The baseline value for a particular variable is generally the last value collected prior to the Cycle 1 Day 1 (C1D1) study visit. For subjects who were randomized but did not receive study treatment, the baseline value is defined as the last value collected prior to randomization. For variables collected on C1D1 without explicit timing, the C1D1 values will be used as the baseline values since the protocol requires these tests or procedures to be performed prior to dosing. Baseline values of neutrophils, hemoglobin and platelets used for response evaluation are the average of all values obtained during screening up to and including the first dosing date (or date of randomization, if not dosed). Screen failure data for re-screened subjects will not be used in analysis; lab data or Bone Marrow data collected prior to the first official screening visit are considered as screen failure data.

7.1 Subject Disposition

Subject disposition including numbers randomized, treated, and treatment discontinuation by reason, as well as the reasons for withdrawal from study will be summarized by treatment group using frequencies and percentages based on information collected on the relevant study case report form pages. The number of subjects screened, screen failures and reasons for failure will be summarized. The number of subjects randomized will also be tabulated by region, country, study site/investigator, and treatment group. The number of subjects in each analysis set will be summarized by treatment group, when applicable. In addition, subject eligibility information including characteristics related to inclusion and exclusion criteria collected during screening will be summarized by treatment group.

Duration of follow-up will be calculated as the date of cutoff - date of randomization and will be summarized using mean, standard deviation, median, minimum and maximum.

7.2 Demographic and Other Baseline Characteristics

The demographic and baseline characteristics include, but are not limited to age, sex, ethnicity, race, geographical region, disease category (MDS vs CMML), time since diagnosis, height, weight, body surface area (BSA), ECOG performance status, number and type of prior cancer therapies, [redacted] risk levels, RBC and platelet transfusion dependences, peripheral blood counts of hemoglobin, neutrophils, platelets, total white blood cells (WBCs) with blasts counts, and bone marrow [BM] blasts counts and presence of baseline genetic mutations.

Time since diagnosis will be calculated as the (date of randomization - date of diagnosis). If the day is missing for date of diagnosis, the 15th of the month is used. If the month is missing, July 1st is used. If the year is missing, the date is left as missing.

Subject demographic and baseline characteristics will be summarized by mean, standard deviation, median, minimum, and maximum for continuous variables; and by counts and percentages for categorical variables. Both the Efficacy and Safety Analysis Sets will be used for the summaries.

7.3 Efficacy Variables and Analyses

Unless otherwise specified, all efficacy analyses will be based on the Efficacy Analysis Set. This section describes the analyses conducted at the primary analysis when approximately 316 death events have occurred within 12 months from the date of the last subject randomization, or after 277 death events have been observed if 316 death events were not reported by 12 months from the date of the last subject randomization. The alpha levels referenced in this section are nominal alpha levels for judging statistical significance, taking into consideration the planned interim analysis and the hierarchical testing order of the (final) primary analysis. The stratification factors used in the analyses will be the randomization stratification factors unless it is necessary to collapse some strata due to analysis difficulties caused by too many strata. The rule for collapsing strata are detailed in Section 7.3.5. If the strata are collapsed for the analysis of the primary endpoint, the collapsed strata will be applied to stratified analyses for all subsequent endpoints.

7.3.1 Primary Efficacy Endpoint and Analyses

OS is the primary efficacy endpoint and is defined as the time, in days from the subject randomization date to the date of death (regardless of cause).

Survival time in days = (earliest date of death or censoring - date of randomization).

In the absence of death at the time of analysis, the survival time will be censored on the last date the subject is known to be alive. Subjects who do not have post-baseline information will be censored at the date of randomization.

OS curves will be estimated using the Kaplan-Meier method and formally compared between the two treatment groups using a 2-sided stratified log-rank test, stratified by the randomization stratification factors.

The null and alternative hypotheses are:

- **Null hypothesis H_0:** Survival curves are the same between the two treatment arms
- **Alternative hypothesis H_1:** Survival curves are different between the two treatment arms

OS will be tested at an overall alpha level of 0.05, as further described in Section 7.3.3. If this primary endpoint reaches statistical significance in favor of guadecitabine at either the interim analysis or final analysis, then the study will be considered positive for efficacy.

The median (and quartiles) duration of OS and the associated 95% CI for each treatment arm will be estimated using the Kaplan-Meier method and the log-log transformation for the survival function.

In addition, the HR and its 95% CI will be estimated using a Cox proportional-hazard model with treatment group as the independent variable and stratified by the same randomization stratification factors as used for the log-rank test. The assumption of proportional hazards will be evaluated.

The following sensitivity analyses will be conducted for evaluating robustness of the treatment effect in OS:

1) The OS will be analyzed using Kaplan-Meier method and compared between the guadecitabine and TC using log-rank test without stratification.
2) An “as-treated” analysis will be conducted for the OS, based on treatment actually received as described in Section 5.2, using the same Kaplan-Meier method and 2-sided stratified log-rank test described above. Subjects who were randomized but not treated will not be included in this analysis.

3) An analysis will be conducted using the same Kaplan-Meier method and 2-sided stratified log-rank test described above. However, the survival time will also be censored on the date the subject receives other anti-leukemia treatments, including chemotherapy but excluding hematopoietic cell transplant (HCT), since HCT are given to subjects whose disease is well controlled and other anti-leukemia therapies are usually given to subjects at the time of their disease progression.

7.3.2 Secondary Efficacy Endpoints and Analyses

Secondary efficacy endpoints include 8-week transfusion independence, marrow CR with transfusion independence, survival rate at 1 and 2 years after randomization, leukemia-free survival, NDAOH, disease response (CR, mCR, PR and HI), time to first response, time to CR, time to best response, duration of complete response, number of RBC or platelet transfusions over the duration of the study treatment, the EQ-5D-5L descriptive system and the EQ VAS.

7.3.2.1 Transfusion Independence

Transfusion independence rate is calculated as the number of subjects with neither RBC nor platelet transfusion for any period of 8 weeks after the initiation of treatment (or C1D1 visit date for subjects randomized to BSC or randomization date for subjects not treated) and up to treatment discontinuation (or 180 days for subjects discontinuing the treatment within 6 months), while maintaining Hgb ≥8 g/dL and platelets $\geq 20 \times 10^9$/L divided by the total number of subjects included in the efficacy analysis. Transfusion dependence at baseline is defined as documentation of 2 or more units of transfusion of RBC or platelet within 56 days of the first dose of study treatment. Transfusion independence rate will be compared between the 2 treatment groups using a Cochran Mantel-Haenszel test stratified by the randomization stratification factors at an overall alpha level of 0.05, as further described in Section 7.3.3. In addition, the Mantel-Haenszel weighted difference in transfusion independence rate between the 2 treatment groups and the associated CI will be provided.

Transfusion independence rates for durations of 16 and 24 weeks will also be presented.

7.3.2.2 Criteria for Response Assessment

The 2006 MDS IWG Criteria (summarized in Table 1) will be used for identifying MDS/CMML subjects with CR, PR, mCR and HI (HI-N, HI-P, HI-E), as well as subjects who did not respond (NR) or whose response status could not be evaluated (NE; information insufficient for response assessment). For the purposes of the planned analyses, the best response will be determined for each subject in the order of CR, PR, mCR, HI, NR and NE. In addition, mCR with HI, which is a subset of mCR subjects who also achieved HI, will also be reported.

Table 1: 2006 MDS IWG Criteria

Complete Response (CR):		
Peripheral: Normal peripheral counts with persistent granulocyte count ≥1.0×10^9/L, platelet ≥100×10^9/L and Hgb ≥11 g/dL. No blasts.		
Marrow: Normal bone marrow with persistent marrow blasts ≤5%. Persistent dysplasia will be noted.		
Partial Response (PR):		
Peripheral: Normal peripheral counts with granulocyte count ≥1.0×10^9/L, platelet count ≥100×10^9/L, and Hgb ≥11 g/dL. No blasts.		
Marrow: Bone marrow blasts >5%, but were reduced by 50% or more from pretreatment levels.		
Marrow Complete Response (mCR):		
Reduction of bone marrow blasts to ≤5% and decrease by 50% or more with or without normalization of peripheral counts.		
Hematological Improvement (HI): lasts at least 8 weeks		
Erythroid Response (HI-E):	Major Response:	Hemoglobin increase ≥1.5 g/dL or relevant reduction of units of RBC transfusions by an absolute number of at least 4 RBC transfusions/8 week compared with the pretreatment transfusion number in the previous 8 weeks.
Platelet Response (HI-P):	Major Response:	Absolute increase of platelet count from ≤20 to >20×10^9/L and by at least 100%, or if more than 20×10^9/L, by an absolute increase of at least 30×10^9/L.
Neutrophil Response (HI-N):	Major Response:	Granulocyte increase ≥100%, and by an absolute increase ≥0.5×10^9/L.

Source: Based on 2006 IWG criteria (adapted from Cheson et al 2006)

7.3.2.3 Marrow CR with Transfusion Independence

The proportion of subjects who achieve mCR (as defined in Table 1) and transfusion independence (as described in Section 7.3.2.1) simultaneously in the same period will be calculated for each group. Subjects who have baseline BM blasts ≤5%, will be included in the denominator for calculation of the response rate. Comparison of the proportions between the two treatment groups will be made using a Cochran Mantel-Haenszel test stratified by the randomization stratification factors, at an overall alpha level of 0.05, as further described in Section 7.3.3. The Mantel-Haenszel weighted difference in the proportions between the 2 treatment groups, and the associated 95% CI will be provided.

7.3.2.4 One Year and Two-Year Survival Rate

One and two-year survival rates are defined as the survival rate at the end of the first and second year from randomization respectively. One and two-year survival rates for each treatment group will be estimated by Kaplan-Meier procedure as addressed in Section 7.3.1. Hypothesis testing will be based on the stratified Kaplan-Meier estimates and standard errors estimated by Greenwood formula using the log-log transformation of the OS function at an overall alpha level of 0.05, as further described in Section 7.3.3. Subjects who do not have death in record will be censored on the last date known to be alive. The stratification factors will be the same as those used in the OS analysis.

7.3.2.5 Leukemia-Free Survival

Leukemia-free survival is defined as the time, in days, from the subject randomization date to the earliest of the dates when a subject has BM or PB blasts ≥20%, conversion to AML or death of any cause. In the absence of either of these events at the time of analysis, leukemia-free survival will be censored on the last date of BM or PB blasts assessment, whichever is later. Subjects who do not have readable BM or PB blast data after randomization will be censored on the date of randomization. Leukemia-free survival curves will be estimated using the Kaplan-Meier method and formally compared between the two treatment groups using a 2-sided log-rank test stratified by the randomization stratification factors. Leukemia-free survival will be tested at an overall alpha level of 0.05, as further described in Section 7.3.3. Medians and quartiles of leukemia-free survival and their 95% CIs will be estimated by the Kaplan-Meier procedure for each treatment group.

7.3.2.6 Number of Days Alive and Out of the Hospital

The date of each hospital admission and discharge will be collected for each subject for at least 6 months (and until disease progression for subjects who have not progressed during the first 6 months or until the subject dies or withdraws consent prior.

The duration of each individual hospital stay in days (regardless of the reason for hospitalization) is calculated as:

Duration of individual hospitalization stay = date of discharge - date of admission

For a subject who is admitted and discharged on the same day the duration of hospital stay will be 0. The total duration of all hospital stays in the first 6 months is the sum of the durations of all individual hospital stays occurring between date of randomization and Day 180. For ease of calculation, one month is defined as 30 days for analyses conducted in this study.

The NDAOH within the first 6 month period is calculated as:

NDAOH6M = 180 - total duration of hospitalization - number of death days before Day 180

For subjects who die within the first 6 month period, the number of death days before Day 180 is calculated as:

Number of death days = (date of Day 180 - date of death).

For subjects who die on or after Day 180 the number of death days before Day 180 will be set to 0.

For subjects who are lost to follow-up within 6 months (expected to be a very small number), the NDAOH6M will be calculated conservatively assuming that the subject would have died the day after the last contact day.

The NDAOH6M will be summarized by treatment group and compared between the 2 treatment groups using an analysis of variance (ANOVA) model at an overall alpha level of 0.05, as further described in Section 7.3.3. The variables used for stratification at randomization will be included in the analysis of variance model as fixed factors.

7.3.2.7 Disease Responses

The best response (CR, PR, mCR, HI, NR, or NE) will be summarized in a table displaying the number and percentage of subjects who meet the criteria for each category. The overall response (OR) rate is calculated as the number of subjects who achieve a best response of CR, mCR, PR or HI divided by the total number of subjects included in the efficacy analysis. The rates of HI-N, HI-P and HI-E, separately as subcategories of HI, will be summarized similarly. These response rates will be summarized for each treatment group along with the corresponding CIs.

In additional, time to first response (CR, PR, mCR or HI), time to CR, and time to best response will be evaluated from date of randomization for responders. Time to first response is defined as the time, in days, from the date of randomization to the first date when any response is achieved. Time to CR is calculated as the time, in days, from the date of randomization to the first date when CR is achieved. Time to best response is similarly defined as the time, in days, from the date of randomization to the first date when a subject's best response, in the order of CR, PR, mCR or HI is achieved. These data will be summarized by treatment group using mean, standard deviation, median, minimum, and maximum.

7.3.2.8 Duration of Complete Response

Duration of complete response (in number of days) will be calculated from the first time a CR is observed to the date of the earliest of the following three events: 1) relapse/disease progression, 2) start of alternative therapy (except HCT) or 3) death. In the absence of any event, the duration of CR will be censored at the last available time point (BM assessment, PB assessment, or safety/long-term follow-up visit) at which an event was not observed. Duration of complete response will be analyzed using a Kaplan-Meier method for subjects who achieved a CR during the study and compared between the two treatment groups using a log-rank test with no stratification. The median and quartiles of duration of complete response, as well as their respective 95% CIs will be provided.

To take the proportion of responders into consideration when analyzing duration of response, a separate analysis including all subjects in the Efficacy Analysis Set will be conducted with a 0 day duration assigned to subjects who did not achieve a CR. The median and quartiles of duration of response, as well as their respective 95% CIs will be provided.

7.3.2.9 Number of RBC or Platelet Transfusions

One RBC transfusion is defined as 1 unit of RBC transfusion. One platelet transfusion is defined as 1 unit of platelet transfusion. Transfusion dates and the number of RBC or platelet transfusions will be collected for each subject for a minimum of 6 months (and until termination of study

treatment if the study treatment lasts for more than 6 months), unless the subject dies or withdraws consent prior.

The total number of RBC transfusions and, separately, the total number of platelet transfusions up to the 6-month time point for each subject is counted from the initiation of treatment (or C1D1 visit date for subjects randomized to BSC or randomization date for subjects not treated) to Day 180. The number of transfusions up to Day 180 will be summarized separately for RBCs and platelets using mean, standard deviation, median and quartiles. The 95% CI of the means will also be provided.

7.3.2.10 EQ-5D-5L Descriptive System Total Score and the EQ VAS

The EQ-5D-5L descriptive system scores and the EQ VAS will be collected for each subject for a minimum 6 months unless the subject dies or withdraws consent prior. The calculation for EQ-5D-5L index value will be performed according to EuroQol group's EQ-5D-5L User Guide (http://www.euroqol.org/about-eq-5d.html). Since there are only a limited number of countries which have rules available for calculating EQ-5D-5L index values, the rules for England (Devlin et al 2016) will be used for all subjects from all countries in this study. These analyses will include only the data collected for each subject during the first 6 months of study participation.

As suggested in the EQ-5D-5L User Guide (http://www.euroqol.org/about-eq-5d.html), the EQ-5D-5L descriptive scores and their dichotomized levels (No problems, Problems) within each EQ-5D dimension (mobility, self-care, usual activity, pain/discomfort, anxiety/depression) will be summarized by time (ie, visit/treatment cycle) descriptively, using counts and proportions. The EQ-5D-5L index value and VAS and their respective changes from baseline will be summarized by time (i.e., visit/treatment cycle) using means, standard deviations, medians and quartiles.

In addition, the changes from baseline (post baseline value - baseline value) of EQ-5D-5L index value, and separately EQ VAS, will be analyzed using a mixed model approach for repeated measures. This model will include the following terms as fixed effects: baseline value, treatment, time, and treatment-by-time interaction. The unstructured covariance matrix will be used to account for the within subject correlation and allow for different variances at different measurement times. The difference of the least squares means between the two treatment groups at each time (ie, visit/treatment cycle) and its corresponding 95% CI will be provided.

7.3.3 Test Sequence and Procedures of Statistical Tests for Efficacy Endpoints

To control the alpha errors associated with testing multiple endpoints, the primary endpoint, OS, will be tested first. If OS reaches statistical significance in favor of guadecitabine, a hierarchical analysis of secondary endpoints will proceed in the following pre-specified order:

1. 8-week transfusion independence.

2. mCR with transfusion independence.

3. Survival rate at 2 years after randomization.

4. Survival rate at 1 year after randomization.

5. Leukemia-free survival.

6. NDAOH during the first 6 months.

A statistically significant test result in favor of guadecitabine for a given endpoint serves as a gatekeeper (Westfall and Krishen 2001) for the testing of statistical significance to proceed to a subsequent secondary endpoint. The overall alpha error rate is controlled at the 0.05 level by following the above testing sequence and procedures.

The efficacy endpoints of disease responses, duration of complete response, number of RBC or platelet transfusions and health-related QOL will be used as supportive evidence of the beneficial treatment effect and will therefore, not be included in the hierarchical testing to control alpha error. Differences of treatment effect for these endpoints and associated 95% CIs, if applicable, will be constructed.

[redacted]

7.3.5 Dealing with Technical Issues Caused by Small Cell Count

If technical issues (such as non-convergence or unstable variance) arise due to small cell counts caused by too many levels of stratification variables, the stratification variables will be collapsed in the following order, until the technical issues have been resolved:

1) Collapse pre-selected TC options (combine IC and BSC).
2) Collapse pre-selected TC options and geographic regions (combine North America and ROW).
3) Collapse pre-selected TC options, geographic regions and collapse disease categories (MDS and CMML).

7.4 Safety Variables and Analyses

Unless otherwise specified, all safety analyses will be performed using the Safety Analysis Set which includes data from all subjects who receive any amount of study treatment (guadecitabine or TC; see Section 5.2). For all safety analyses, summaries by treatment group (guadecitabine and TC) will be provided [redacted] Data will be summarized by guadecitabine and each TC option within a given preselected TC option.

Safety is assessed by subject-reported and investigator-observed AEs, and 30-day, 60-day and 90-day all-cause mortality, along with concomitant medications, physical examination, clinical laboratory tests (hematology, serum chemistry, and urinalysis), vital signs, ECOG performance status, and ECGs. Safety is also assessed by exposure to guadecitabine or TC, reasons for discontinuation, deaths and causes of deaths.

All safety data collected during the study will be included in the study database. All safety data collected during the study will be used for generation of safety summary tables, with the exception of AEs and medications. The AE and medication summary tables will only include treatment-emergent AEs and concomitant medications as defined in Sections 7.4.2 and 7.4.4.

7.4.1 Study Treatment Exposure

Cycle 1 Day 1 is defined as the first day of study treatment after randomization; cycle days are counted sequentially thereafter. Cycle 2 Day 1 is the first day of Cycle 2 regardless of treatment

delays. The designated cycle duration is 28 days but any cycle could be prolonged to >28 days to allow blood count recovery if deemed clinically necessary. This convention for determining the start and stop dates for cycles is maintained until treatment is permanently discontinued. For easy presentation, Cycle x Day y is often abbreviated as CxDy in this document and statistical outputs.

Frequency counts and percentages of dose cycles received (or CxD1 visits completed for subjects randomized to BSC), dose cycles delayed, and dose reduced cycles as well as the reason for reduction and delays, when available, will be summarized by treatment group. These summaries will be provided at both subject and cycle levels, and will be based on dose administrations, dose reductions and dose delays identified by the study site and collected on the dosing CRFs. A dose reduced cycle is defined as a cycle where the study site has indicated that the dose was reduced or in which the planned dose for Cycle 2 and above was lower than the planned Cycle 1 dose or the number of doses taken with the cycle is less than the planned number of doses. Percentage of intended dose is equal to actual total dose divided by planned total dose calculated for each cycle. Dose intensity, presented as the incidence of subjects receiving <80% of their intended dose will be summarized by cycle for the Guadecitabine treatment group. Both completed or partially completed dose cycles are counted in this summary. For BSC, dose delay, reduction, and percentage of intended dose will not be summarized.

7.4.2 Adverse Events

AE terms reported by study subjects or observed by investigators will be mapped to the appropriate System Organ Class (SOC) and Preferred Term (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA). Severity of AE will be graded using Common Terminology Criteria for Adverse Events (CTCAE) version 4.03.

Treatment-emergent AEs are defined as events that first occurred or worsened after the first dose of study drug given on C1D1 (or the C1D1 visit date for subjects randomized to BSC) until 30 days after the last dose of study treatment (or until the Safety Follow-up visit date for subjects randomized to BSC), or the start of an alternative anti-leukemia or MDS treatment, whichever occurs first, with the following exceptions: events that occurred after 30 days beyond the last dose of study treatment (or after the Safety Follow-up visit date for subjects randomized to BSC), or the start of an alternative anti-leukemia or MDS treatment will also be considered treatment-emergent if the events are both serious and related to the study treatment. For the purpose of determining whether an AE is a treatment-emergent AE, incomplete AE start and stop dates will be imputed conservatively.

All AE data collected in the study database will be listed, including those that are not treatment emergent. However, safety tables will be generated based only on treatment-emergent AEs.

An overall safety summary table containing counts and percentages of subjects with any AE, any AE Grade ≥3, AE leading to treatment discontinuation, AE leading to drug interruption, AE leading to dose reduction, any serious AE (SAE), and subcategories of SAEs (fatal vs non-fatal) will be produced by treatment group. A similar table with related AE counts will also be produced

by treatment group. Related events are those that the investigator considered to be suspected to be related to study treatment as described in the study protocol.

The number and percentage of subjects experiencing AEs will be summarized by MedDRA SOCs (sorted alphabetically) with PTs sorted by decreasing frequency within each SOC, and by CTCAE grade. The number and percentage of subjects experiencing AEs will also be summarized by PT and sorted by event frequency. Related AEs, AEs Grade ≥3, related AEs Grade ≥3, serious AEs, related serious AEs, AEs with an outcome of death, and AEs leading to treatment discontinuation, AEs leading to drug interruption and AEs leading to dose reduction will be summarized similarly. In summarizing AEs, if a subject reports the occurrence of a particular AE more than once, the event is only counted once with its worst CTCAE grade.

7.4.3 30- and 60-Day All-cause Mortality

The 30- and 60-day all-cause mortality will be calculated as the proportions of subjects who have died within 30 or 60 days of study treatment initiation based on each subject's date of death relative to C1D1 (ie, date of death minus date of C1D1). Subjects who die within 30 days will also be included in the 60-day mortality calculations. To avoid using a different denominator for calculation of the percentages for 30- and 60-day mortality, subjects who were lost-to-follow-up within 30- or 60-days from C1D1 (assumed to be a very small number, if any) will be considered alive for the corresponding 30- and 60-day mortality calculations. Ninety-day all-cause mortality will be calculated similarly.

In addition, a summary of all deaths, by cause of death (AE, progressive disease, other) and treatment group will be provided for the Efficacy Analysis Set (all randomized subjects).

7.4.4 Concomitant Medications

Medications will be coded using the WHO Drug Dictionary.

Concomitant medications are the medications taken with a start date on or after the start of the administration of study treatment or the C1D1 visit date for subjects randomized to BSC), or those with a start date before C1D1 and a stop date on or after C1D1. Medications taken beyond 30 days from the last dose of study treatment (or after the Safety Follow-up visit date for subjects randomized to BSC), or after the start of an alternative anti-leukemia or MDS treatment are not considered concomitant medications, unless they are used for treating a related SAE.

For the purpose of determining whether a medication is a concomitant medication, incomplete medication start and stop dates will be imputed conservatively.

Concomitant medications will be summarized by WHO Drug Dictionary Therapeutic Subgroup (ATC level 2) and Drug Name, sorted alphabetically, using counts and percentages.

Special interest concomitant medications include anti-emetic drugs, growth factors (including G-CSF, GM-CSF and ESAs), anti-infective medications (including but not limited to anti-bacterials,

anti-mycotics, anti-mycobacterials, anti-virals and immunoglobulins) and hydroxyurea given to reduce high counts during study treatment and not as part of a subsequent anti-leukemia treatment. These concomitant medications will be tabulated separately. Transfusions will be described separately as part of the efficacy analyses (Section 7.3.2.9).

7.4.5 Laboratory Tests

Data from different local laboratories will be standardized to consistent SI units, and presented in data listings. Laboratory values recorded as an interval such as "≥x", "<x", or "2+" will be handled, if necessary for calculation purposes, following the data programming standards as detailed in the study's Analysis Data Model (ADaM) specifications.

Laboratory values will be graded, if relevant and possible, by CTCAE version 4.03 in conjunction with the Harrison (18th edition) lab book normal values (Kratz et al 2012). Shift tables will display (1) shift from baseline grade to the worst grade during the study, and (2) shift from baseline grade to the last grade at the end of study.

Summaries will also be provided of the incidence of all new or worsening laboratory abnormalities (any CTCAE grade) and new or worsening CTCAE Grade ≥3 laboratory abnormalities by parameter. A listing of potential Hy's Law cases will be provided. In addition, for selected laboratory parameters, a figure of mean values by visit will be generated.

7.4.6 Vital Signs

Vital signs will be summarized by visit using the proportion of subjects who have vital sign values too high or too low, according to the conventionally accepted vital sign normal ranges as listed below:

- Pulse rate ≥110 bpm.
- Pulse rate ≤50 bpm.
- Diastolic blood pressure ≥110 mmHg.
- Diastolic blood pressure ≤55 mmHg.
- Systolic blood pressure ≥180 mmHg.
- Systolic blood pressure ≤80 mmHg.
- Respiration rate ≥20 breaths/min.
- Body temperature ≥39°C.

7.4.7 Electrocardiogram

At each ECG assessment time point (pre- and post-dose on Day 1 of Cycle 1 and at the safety follow-up visit), the mean of the available triplicate electrocardiogram values will be entered in the CRF. For the ECG parameters (heart rate, PR interval, QRS duration, QT interval and QTcF),

the value and respective changes from baseline will be summarized by visit using mean, standard deviation, median, minimum and maximum. Summaries will also be provided of the maximum post-baseline absolute QTcF and the maximum post-baseline increase in QTcF.

7.4.8 ECOG Performance Status

ECOG performance status will be summarized by visit, at all scheduled visits where performance status was assessed, using counts and percentages.

7.4.1 Physical Examination

Relevant physical examination data may be presented in a data listing.

[redacted]

7.7 Interim Analyses and Data Monitoring

Data are reviewed by an independent DMC at regular intervals primarily to evaluate safety during study conduct. The committee operates independently from the Sponsor and the clinical investigators.

One formal interim analysis of OS was planned in the protocol with a maximum spendable alpha (2-sided) of 0.01using the Lan-DeMets (Lan and DeMets 1983) implementation of the O'Brien-Fleming (O'Brien and Fleming 1979) boundary adjusted by the actual proportion of events at the

interim relative to the target final 316 death events. The interim analysis was conducted by the DMC at one of its scheduled meetings on 16 April 2019 after approximately half (ie, approximately 158) of the required death events had occurred. All data available at the time of the interim analysis was included in the interim analysis. The nominal two-sided alpha values for the interim analysis of OS at 50% information time point is 0.00014. The DMC was guided by this monitoring boundary in the development of their recommendations regarding whether the trial should continue. If the P-value at interim was greater than 0.00014, the nominal alpha for the final analysis will be 0.04998. The actual nominal alpha value to be used in the final analysis is dependent on the actual alpha spent in the interim analysis.

The alpha to be spent at the interim time point for the secondary endpoints was the same as that of the primary endpoint (α_{interim}). The nominal alpha at the final analysis time point for the secondary endpoints will be different from the final nominal alpha of the primary analysis (α_{final}) due to different fractions of information accumulated at the time of interim analysis. For simplicity, the nominal two-sided alpha at the final analysis time point for all the secondary endpoints involved in hierarchical testing will be 0.05 - α_{interim}. The statistical tests at each of the interim and final analysis time points for secondary endpoints will be performed in the order of 8-week transfusion independence, mCR with transfusion independence, 2-year survival rate, 1-year survival rate, leukemia-free survival and NDAOH based on the nominal alpha as described above. The test process stops when it reaches the first non-significant result or when all the tests listed in the hierarchical testing have been completed. The overall type 1 error is controlled at the two-sided 0.05 level.

The boundary for lack of benefit used by the DMC in the added futility analysis at one of the scheduled DMC meetings on 12 September 2018 when ~100 death events had occurred was the O'Brien-Fleming 'lower boundary' that rules out the hazard ratio of 0.802 (one-sided alpha of 0.025). With such a boundary, there would be in essence no power loss due to the added interim futility analysis and therefore no sample size adjustment was necessary.

7.8 Handling of Missing Data and Other Data Anomalies

No missing data imputations are planned for the study, except as specified. Subjects lost to follow-up will be included in statistical analyses to the point of their last evaluation.

7.9 Handling of Protocol Deviations

Protocol deviations that occur during the study are captured by study monitors and recorded in the CRO's clinical trial management system. Study medical monitors conduct regular reviews of all recorded protocol deviations to ensure the quality conduct of the study. Study medical monitors also identify and categorize important protocol deviations. Important protocol deviations will be summarized by deviation category using counts and percentages. A data listing of all protocol deviations will also be provided.

Protocol deviations related to the COVID-19 pandemic will be noted in the system and an additional listing of protocol deviations related to COVID-19 will be generated.

8.0 REFERENCES

Bejar R, Steensma DP. Recent developments in myelodysplastic syndromes. Blood. 2014;124(18):2793-2803.

Bennett JM, Catovsky D, Daniel MT, et al. Proposals for the classification of the myelodysplastic syndromes. Br J Haematol. 1982;51(2):189–199.

Cheson BD, Bennett JM, Kantarjian H, et al. Report of an international working group to standardize response criteria for myelodysplastic syndromes. Blood. 2000;96(12):3671-3674.

Cheson, BD, Greenberg PL, Bennett JM, et al. Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. Blood. 2006;108(2):419–425.

Devlin, N, Shah, K, Feng, Y, Mulhern, B and Van Hout, B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Research Paper 2016;16/01.

Fenaux P, Mufti G, Hellstrom-Lindberg E, et al. 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. 2009;10(3):223–232.

Heaney ML, Golde DW. Myelodysplasia. N Engl J Med. 1999;340(21):1649–1660.

Issa J, Roboz G, Rizzieri D, et al. Safety and tolerability of guadecitabine (SGI-110) in myelodysplastic syndrome and acute myeloid leukemia: a multicentre, randomised, dose-escalation phase 1 study. Lancet Oncol. 2015;16(9):1099–1110.

Kratz A, Pesce MA, Basner RC, Einstein AJ. Laboratory Values of Clinical Importance (Appendix). In: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J editors. Harrison's principles of internal medicine, 18th edition. New York: McGraw-Hill Professional. 2012.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-663.

Meggendorfer M, Roller A, Haferlach T, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). Blood. 2012;120(15):3080–3088.

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549-56.

Padron E, Komrokji R, List AF. The clinical management of chronic myelomonocytic leukemia. Clin Advan Hematol Oncol. 2014;12(3):172–178.

Seymour JF, Fenaux P, Silverman LR, et al. Effects of azacitidine compared with conventional care regimens in elderly (≥75 years) patients with higher-risk myelodysplastic syndromes. Crit Rev Oncol Hematol. 2010;76(3):218–227.

Silverman LR. The myelodysplastic syndrome. In: Kufe DW, Pollock RE, Weichselbaum RR, et al, editors. Holland-Frei Cancer Medicine. 6th ed. Hamilton (ON): BC Decker; 2003. Chapter 126.

Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. J Stat Plan Inference. 2001;99(1):25–40.