Official Title: A Multicenter, Double-Blind, Randomized, Placebo-Controlled, Phase III Study of Idasanutlin, an MDM2 Antagonist, With Cytarabine Versus Cytarabine Plus Placebo in Patients With Relapsed or Refractory Acute Myeloid Leukemia (AML)

NCT Number: NCT02545283

Document Date: SAP Version 3: 30-September-2019

STATISTICAL ANALYSIS PLAN

TITLE: **A MULTICENTER, DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED, PHASE III STUDY OF IDASANUTLIN, AN MDM2 ANTAGONIST, WITH CYTARABINE VERSUS CYTARABINE PLUS PLACEBO IN PATIENTS WITH RELAPSED OR REFRACTORY ACUTE MYELOID LEUKEMIA (AML)**

PROTOCOL NUMBER:	WO29519
STUDY DRUG:	Idasanutlin (RO5503781)
VERSION NUMBER:	3
EUDRACT NUMBER:	2014-003065-15
SPONSOR:	F. Hoffmann-La Roche Ltd
PLAN PREPARED BY:	[redacted], Ph.D. [redacted], M.Sc. [redacted], M.Sc.
DATE FINAL:	25 September 2019

STATISTICAL ANALYSIS PLAN APPROVAL

Date and Time(UTC)	Reason for Signing	Name
30-Sep-2019 14:40:49	Company Signatory	[redacted]

CONFIDENTIAL

RATIONALE FOR STATISTICAL ANALYSIS PLAN VERSION 3

The purpose of this draft statistical analysis plan amendment (Version 3) is to provide updates accompanying protocol version 6 as follows:

- To add an interim analysis for OS at an information fraction of 80%.
- To update the primary/secondary analysis population to only TP53 wild type (WT) Intent-to-treat (ITT) patients (all ITT patients are only assessed in exploratory endpoints), and to add the definition of the TP53 WT ITT primary population.
- To redefine list and order of hierarchical testing framework for secondary endpoints
- To add exploratory endpoints
- To add sensitivity analyses
- To make handling of missing data more precise

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

1. BACKGROUND

This document is based on the statistical section of the study protocol and will provide more details to the planned statistical analyses. Any analyses that deviate from or are added to those outlined in the protocol will be delineated in this document.

Study WO29519 was developed to examine the efficacy and safety of idasanutlin (RO5503781) in combination with cytarabine (idasanutlin + cytarabine) in patients with relapsed or refractory acute myeloid leukemia (AML) compared with placebo in combination with cytarabine (placebo + cytarabine).

2. STUDY DESIGN

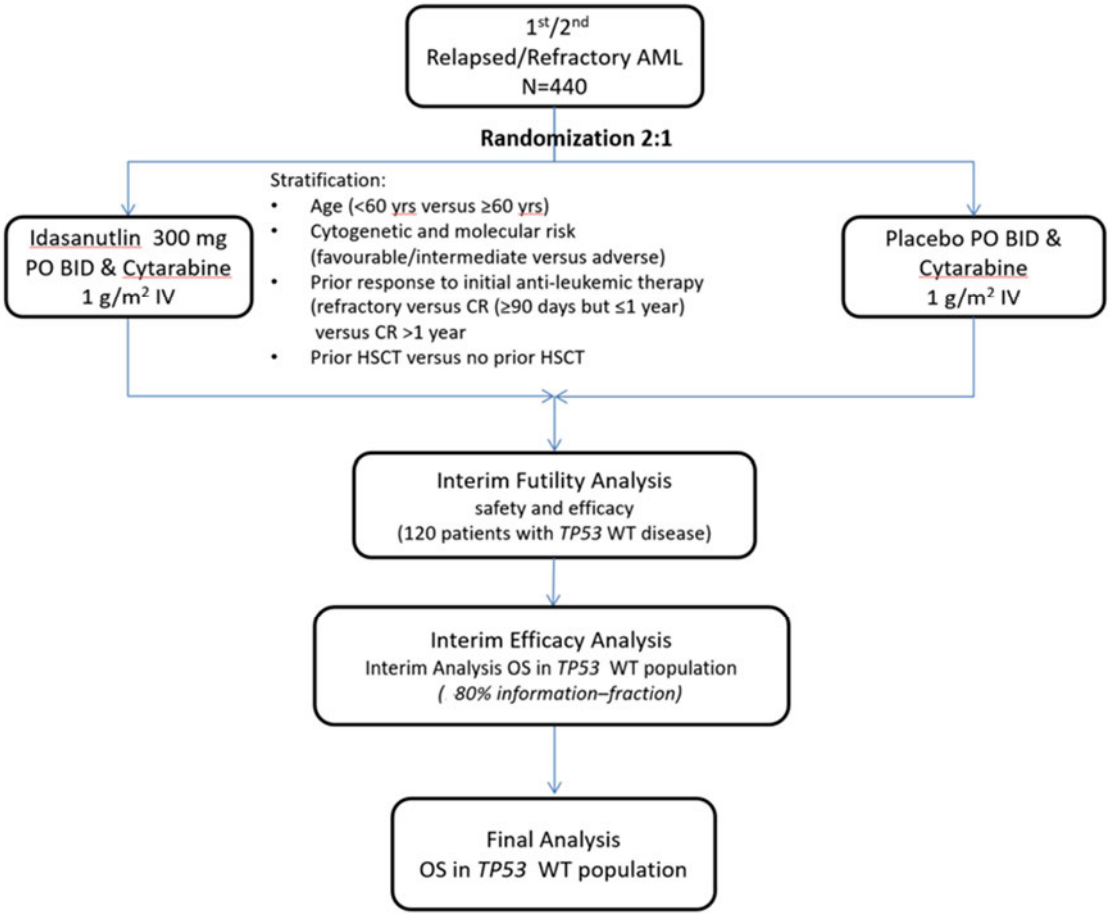
This is a Phase III, multicenter, double-blind, randomized, placebo-controlled study comparing the efficacy and safety of idasanutlin + cytarabine versus placebo + cytarabine in patients with relapsed or refractory AML.

Approximately 440 patients with AML who have relapsed following or are refractory to cytarabine-containing standard induction chemotherapy after one or two prior cytarabine-containing induction chemotherapy regimen(s) are planned to be enrolled (see study schema in Figure 1).

Patients will be enrolled from approximately 80 centers in 18 countries across Europe, the Americas, and the Asia-Pacific region. Patients will be randomized in a 2:1 ratio to receive either idasanutlin + cytarabine (i.e., the experimental treatment arm) or placebo + cytarabine (i.e., the control treatment arm). The randomization will be stratified by age (<60 years old vs. ≥ 60 years old), cytogenetic and molecular risk (favorable/intermediate vs. adverse) at initial diagnosis, prior response to initial anti-leukemic therapy (refractory vs. CR ≥ 90 days but ≤ 1 year vs. CR > 1 year), and prior autologous or allogeneic hematopoietic stem cell transplant (HSCT) versus no prior HSCT.

Patients will receive induction treatment during Cycle 1 with 300 mg idasanutlin or placebo twice a day and 1 g/m^2 cytarabine for 5 days followed by 23 days of rest. If needed, an additional 4 weeks up to Day 56 can be used to allow for blood count recovery. For patients who achieve clinical response after induction and for whom HSCT for consolidation is not an option, it is strongly recommended to continue study medication with a maximum of two additional cycles of consolidation (Cycles 2 and 3) unless clinically contraindicated. Patients who have a good partial response (PR) at the end of Cycle 1, can receive another cycle of induction upon Sponsor's approval, and their responses will not be taken into account for the interim futility analysis under Version 3 of the protocol.

Figure 1 Study Schema



AML = acute myeloid leukemia; BID = twice daily; CR = complete remission; HSCT = hematopoietic stem cell transplant; IV = intravenous; OS = overall survival; PO = orally; WT = wild type

Idasanutlin interferes with the binding of MDM2 to tumor protein -53 (TP53), thereby stabilizing p53 and allowing it to exert its anti-tumor activity. Because not all *TP53* mutations lead to a loss-of-function of the p53 protein, some patients with mutant *TP53* may still respond to treatment. The study will therefore allow enrolment of all patients irrespective of *TP53* mutation status. However, to best assess idasanutlin efficacy in AML, the primary analysis will be tested in *TP53* WT patients (see definition in section 4.1.1.1 and Appendix 2), representing about 85% of all patients enrolled. It is expected that about 15% of the enrolled patients in this study will carry a *TP53* mutation.

Prior to the final analysis, a non-binding interim analysis for futility is planned as well as a binding interim analysis for efficacy. Both interim analyses will be conducted by an iDMC (independent Data Monitoring Committee).

The futility interim analysis will be based on the endpoints of confirmed CR + CR with incomplete platelet recovery (CRp), event-free survival (EFS), and other safety data (see details of data monitoring in Section 3.2).

The interim analysis for efficacy on OS is planned at an information fraction of 80% OS events (220 OS events) in TP53 WT patients (see Section 4.11.3).

Accompanying this efficacy-interim-analysis, a non-binding futility assessment will be performed based on an OS hazard ratio (HR) greater than or equal to one.

With regard to study continuation on the basis of OS data at the interim efficacy analysis, a final OS analysis would occur at 275 OS events in *TP53 WT* patients and a follow up OS analysis will occur at end of study.

2.1 PROTOCOL SYNOPSIS

The protocol synopsis is in

2.2 OUTCOME MEASURES

2.2.1 Primary Efficacy Outcome Measures

The primary endpoint for this study is OS in the *TP53* WT population. OS is defined as the time from randomization to death from any cause. OS for patients who have not died at the time of the analysis will be censored at the last date the patient was known to be alive.

2.2.2 Secondary Efficacy Outcome Measures

The secondary efficacy outcome measures for this study are as follows, for the *TP53* WT population:

- CR is defined as *a* complete remission as reported by the investigator *over* the induction cycle (Cycle 1).
- EFS is defined as the time from the date of randomization to treatment failure (failure to achieve CR, set as day of final response assessment), relapse from CR, or death from any cause, whichever occurs first. Patients with no EFS event will be censored at the date of last response assessment (pre-HSCT) or last date known to be alive (post-HSCT). Patients with no EFS event and no post-baseline response assessment will be censored at date of randomization.
- Overall remission (ORR; CR, CRp, and complete remission with incomplete blood count recovery [CRi]) over the induction cycle (Cycle 1)
- DOR (or Duration of CR) is defined only for patients who achieve a CR and is measured from the date of achievement of CR until the date of relapse from CR, or death from any cause. Patients who do not have a DOR event will be

censored the date of last response assessment (pre-HSCT) or last known alive date (post-HSCT).

- The proportion of HSCT following a CR is defined as proportion of patients undergoing allogeneic HSCT following the corresponding CR
- CR in clinically actionable mutation defined subpopulations (FLT3, IDH1 and IDH2)
- OS in clinically actionable mutation defined subpopulations (FLT3, IDH2 and IDH2)

2.2.3 Safety Outcome Measures

The safety and tolerability of idasanutlin + cytarabine in comparison with placebo + cytarabine will be assessed on both the *TP53* WT safety population and the all-patient safety population using the following safety outcome measures:

- Incidence, nature, severity, and seriousness of adverse events
- Incidence of clinically significant laboratory abnormalities
- Electrocardiograms (ECGs)
- Vital signs
- Death, including 30- and 60-day mortality rates

2.2.4 Exploratory Outcome Measures

The pharmacodynamic (PD)/biomarker outcome measures for this study are as follows:

- PD biomarkers and other biomarkers (e.g., protein, nucleic acid, and other tumor cell-derived markers related to the mechanism of action of the study drug, which include, but are not limited to, markers of MDM2 or p53 pathway alterations) that may help predict patient subpopulations who are more likely to respond to the therapies in this study.
- Minimal residual disease (MRD)
    - Kinetics of MRD by flow cytometry and association of MRD response with outcomes
    - MRD by tracking AML-specific mutations and translocations and/ or expression of AML-related genes such as *Wilms tumor 1* [*WT1*])
- Prognostic value of disease related mutations
- Serum macrophage inhibitory cytokine-1 (MIC-1) profile (raw and/or adjusted from baseline as a percentage of change)
- 4-gene signatures by quantitative real-time polymerase chain reaction (qRT-PCR) as well as gene expression levels of the individual components of the signature (including MDM2)

- MDM2 protein expression measured by flow analysis

The efficacy exploratory Outcome Measures are as follows:

- Rate of CR_{MRD} (CR without MRD as per European LeukemiaNet [ELN] guidelines [Dohner 2017] at a 0.1% cut-off for negativity as per ELN MRD working party consensus [Schuurhuis 2018]) within the ITT-all-patient and TP53 WT populations
- OS within the ITT-all-patient population
- Secondary efficacy endpoints within the ITT-all-patient population.
- CR during treatment: defined as complete remission as reported by the investigator at any time during treatment
- The proportion of HSCT as proportion of patients undergoing allogeneic HSCT regardless of response
- The proportion of HSCT following a CR/CRp/CRi is defined as proportion of patients undergoing allogeneic HSCT following the corresponding response (i.e. CR or CRp or Cri)
- Leukemia-free Survival (LFS, or duration of OR) is defined only for patients who achieve a remission (CR, CRp, or CRi) and is measured from the date of achievement of remission until the date of relapse from CR, CRp, CRi, or death from any cause. Patients who do not have an LFS event will be censored at the date of last response assessment.

2.2.5 Pharmacokinetic Outcome Measures

The pharmacokinetic (PK) outcome measures for this study are as follows:

- Apparent clearance and apparent volume of distribution as well as the maximum concentration observed in plasma (C_{max}), steady-state concentration at the end of a dosing interval (i.e., just prior to next drug administration [C_{trough}]), area under the concentration–time curve during one dosing interval, area under the concentration–time curve from 0 to 24 hours, and half-life of idasanutlin (and M4 metabolite RO6802287)
- Total clearance and volume of distribution of cytarabine
- Effect of idasanutlin on cytarabine PK
- Effect of cytarabine on idasanutlin PK

The PK outcome measures will be investigated by the Roche Modeling and Simulation group. Results of the PK analyses will be reported in the population PK report.

2.2.6 Patient-Reported Outcome Measures

The patient-reported outcome (PRO) measures for this study are as follows:

- European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30)

- EuroQol 5-Dimension 5-Level (EQ-5D-5L) questionnaire

2.3 DETERMINATION OF SAMPLE SIZE

A mechanistic simulation model was used to determine sample size based on the following assumptions:

- Final analysis on OS in patients with *TP53* wild type (WT) using a 2-sided log-rank test at the 5% significance level
- 85% power to detect an OS HR (HR) of 0.67 for idasanutlin + cytarabine versus placebo + cytarabine in patients with *TP53* WT, corresponding to a 33% risk reduction of death, corresponding to an improvement in median OS from 6 months to 9 months (50%)
- Proportions of long-term survivors of 8.0% in the placebo + cytarabine treatment arm and 16.1% in the idasanutlin + cytarabine treatment arm

To compute the necessary number of events, OS times were simulated based on the following assumptions:

- All simulated OS times for patients not considered to be long-term survivors were exponentially distributed.
- Probability of being a CR in the placebo + cytarabine treatment arm was 0.16.
- Probability of being a CR in the idasanutlin + cytarabine treatment arm was 0.323, implying an odds ratio for CR of 2.5 for idasanutlin + cytarabine versus placebo + cytarabine.
- Probability for a CR to be a long-term survivor was 0.5 in either treatment arm.
- An annual dropout rate was 5%.

Based on these assumptions, approximately 275 OS events in patients with *TP53* WT are required to provide 85% power for the primary analysis. The minimum detectable OS HR in this study is 0.78, assuming 275 OS events and a significance level of 5%, which corresponds to a minimal detectable median improvement from 6 to 7.7 months assuming exponentiality.

All patients will be randomized in this study. Assuming 85% of patients will have *TP53* WT and 15% will have *TP53* mutant disease, approximately 440 patients will be enrolled over approximately 29 months, corresponding to an estimated number of 374 patients with *TP53* WT.

In Version 6 of the protocol, an interim analysis for efficacy on OS of *TP53* WT patients was added. The sample size assumptions above remain unchanged other than the recruitment time (now estimated as 41 months). The interim analysis will occur at an information-fraction of 80% providing a power of 83% for the final OS analysis (see section 4.11.3).

2.4 ANALYSIS TIMING

The final OS analysis will occur when approximately 275 OS events in patients with *TP53* WT have been observed. It is expected that this analysis will take place approximately 4 years after enrollment start (assuming 9 *TP53* WT patients enrolled per month).

A follow up (updated) OS analysis will be performed at the end of the study, expected approximately 5.5 years after enrollment start (2 years after the last patient is enrolled or after all patients have died, whichever occurs first).

A non-binding futility interim analysis will be performed by the independent data coordinating center (iDCC) and reviewed by the independent Data Monitoring Committee (iDMC). The futility interim analysis will be based on the confirmed CR response, which is defined in Section 4.4.2.

At the time of the study design, the timing of the futility and final analyses have been estimated assuming that on average 12 patients per month are enrolled for the first 15 months, followed by 18 patients per month for 14 months and 8 patients for the last month.

It was expected that the futility interim analysis will occur approximately 18 months after study start. Further details on the interim analyses, such as stopping rules and criteria, are provided in the iDMC charter and in section 4.11.

An interim OS analysis will occur at an information-fraction of 80% (see Section 4.11.3).

3. STUDY CONDUCT

3.1 RANDOMIZATION ISSUES

Approximately 440 patients will be enrolled in this 2:1 randomized trial, corresponding to a higher probability to be in the idasanutlin+cytarabine treatment arm (i.e., about 293 patients in the idasanutlin+cytarabine treatment arm and 147 patients in the placebo+cytarabine treatment arm).

Randomization will be performed through the interactive web response system (IWRS) using a stratified permuted block randomization scheme. The randomization stratification factors are as follows:

- Age ($<$ 60 years vs. $\geq$ 60 years)
- Cytogenetic and molecular risk (favorable/intermediate vs. adverse) at initial diagnosis
- Prior response to initial anti-leukemic therapy (refractory vs. CR $\geq$ 90 days but $\leq$ 1 year vs. CR $>$ 1 year)

- Prior HSCT versus no prior HSCT

Discordance between the IWRS and the electronic Case Report Form (eCRF) data will be summarized.

This is a double-blind study (i.e., neither the sponsor personnel, the patients, nor the treating physician will know the assigned treatment arm). However, the randomization code will be made available to the bioanalytical manager to facilitate the analysis of PK samples.

3.2 DATA MONITORING

An iDMC will be used to assess patient safety and treatment efficacy at defined intervals. Specific details of the schedule of iDMC activities can be found in the iDMC charter.

The iDMC will first meet to assess the safety of the study after 30 patients have been enrolled and completed the first cycle with clinical response assessment available. Thereafter, the iDMC will have regular safety review meetings approximately every 3 months until the futility interim analysis. After this interim futility analysis, the iDMC will meet approximately every 6 months to assess the safety of the study. In addition, the iDMC will perform an interim efficacy analysis on OS at an information-fraction of 80% (see section 4.11.3). Both the Sponsor and the iDMC can request ad hoc iDMC meetings for any reason.

Following each meeting, the iDMC will recommend to the Sponsor whether the study should continue according to the protocol or may suggest changes to the protocol based on the outcome of data review.

At the futility interim analysis, the iDMC will also perform an ordinary safety review and a review of biomarker data. The iDMC will make recommendations on the following:

- Safety monitoring (including gastrointestinal toxicity and early death criteria)
- ECG monitoring
- Biomarker classification (see Section 4.5)

The details about the iDMC recommendation after the futility interim analysis are described in Section 4.11.1, each safety analysis in section 4.11.2 and the interim efficacy analysis in section 4.11.3.

The Sponsor will make the final decision for continuation or discontinuation of the study, or early unblinding, on the basis of the iDMC's recommendation (see Section 6.4 of the iDMC Charter).

4. STATISTICAL METHODS

Except where specified otherwise, continuous data will be summarized with descriptive statistics (number of non-missing data, mean, standard deviation, median, minimum, and maximum). Categorical parameters data will be described using frequencies and percentages.

Descriptive summaries will be provided by treatment group. The main treatment comparison is between idasanutlin + cytarabine and placebo + cytarabine in *TP53* WT patients and all patients who have been randomized. All the statistical tests will be 2-sided.

4.1 ANALYSIS POPULATIONS

4.1.1 Randomized Population

4.1.1.1 *TP53* Wild type Intent-to-Treat Primary Population

All randomized *TP53* WT patients will be included in the *TP53* intent-to-treat (*TP53* ITT) population. The *TP53* ITT population is the primary population for the analysis of the primary and secondary efficacy endpoints (see 4.4). Efficacy analyses will be done according to the randomized treatment arm regardless of which treatments the patients received.

Patients with wild-type *TP53* and *TP53* mutants with retained functionality will, together, form the *TP53* WT ITT primary population based on the variant types as described in Appendix 2.

The futility interim analysis population will comprise the first 120 patients with *TP53* WT who were randomized, have received at least 80% of first treatment cycle (of the total planned dose of each study compound) and have data for response (including confirmed response for those with CR, CRp at end of Cycle 1) or who were withdrawn before time of response/confirmation.

4.1.1.2 All-Patient Intent-to-Treat Population

All randomized patients, that is, patients who are *TP53* WT or mutant or with a mutation status at screening unknown, will be included in the intent-to-treat (ITT) population. Efficacy analyses will be done according to the randomized treatment arm regardless of which treatments the patients received.

4.1.2 Per Protocol Population

A per protocol analysis population is not defined for this study. Major protocol deviations will be summarized by treatment arm.

4.1.3 Pharmacokinetic-Evaluable Population

The PK-evaluable population includes all patients with PK data. PK analyses will be performed on the PK-evaluable population according to the treatment the patients received.

4.1.4 Safety Population

The all-patient safety population will include all patients who received any amount of study drug (i.e., idasanutlin, placebo, or cytarabine) and the safety *TP53* WT population will include all patients with TP53 WT and TP53 mutants with retained functionality based on the variant types as described in Appendix 2 who have received any amount of study drug. Patients will be analyzed according to the received treatment, that is, patients who received at least one dose of idasanutlin for whatever reason will be analyzed under the idasanutlin + cytarabine treatment arm. Patients who only received cytarabine and/or placebo will be analyzed under the placebo + cytarabine treatment arm.

4.2 ANALYSIS OF STUDY CONDUCT

The following analyses of study conduct will be provided by treatment arm for all randomized patients unless specified differently:

- A summary of the randomized patients by treatment arm, center, and country
- A summary of the analysis populations
- A summary of the patient disposition, study treatment administration, and reasons for study treatment discontinuation
- A summary of observation time
- A summary of the major protocol deviations

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Descriptive summaries will be provided by treatment arm. These will include, but are not limited to, the following parameters:

- Stratification factors
  - Age (< 60 vs. ≥ 60 years)
  - Cytogenetic and molecular risk (favorable/intermediate vs. adverse)
  - Prior response to initial anti-leukemic therapy (refractory vs. CR ≥ 3 months but ≤ 1 year vs. CR > 1 year)
  - Prior HSCT versus no prior HSCT
- Demographic characteristics at baseline:
  - Gender
  - Age (years)
  - Race

Eastern Cooperative Oncology Group performance status

- AML disease characteristics at baseline

  AML disease status at study entry (first relapse, second relapse, first refractory to induction, or second refractory to induction)

  Number of prior treatment regimens

  European LeukemiaNet Risk category at initial diagnosis

  TP53 status

  Type of previous HSCT (allogenic vs. autologous)

Should country policy prohibit the collection of race and/or age information, patients will appear in the missing category of summary tables.

4.4 EFFICACY ANALYSIS

The primary and secondary efficacy outcomes will be assessed for the *TP53* WT ITT population (primary population), as defined in Section 4.1.1.1.

4.4.1 Primary Efficacy Endpoint

The OS analysis will assess the null hypothesis of equality of OS distributions in the idasanutlin + cytarabine and placebo + cytarabine treatment arms in *TP53* WT patients:

H_0: $OS_{idasanutlin+cytarabine} = OS_{placebo+cytarabine}$ versus H_1: $OS_{idasanutlin+cytarabine} \neq OS_{placebo+cytarabine}$

OS distributions will be compared between the treatment groups using a 2-sided stratified log-rank test at an overall significance α level of maximum 4%. The stratification factors will be the randomization stratification factors: age (<60 years vs. ≥ 60 years), cytogenetic and molecular risk (favorable/intermediate vs. adverse), prior response to initial anti-leukemic therapy (refractory vs. CR ≥ 3 months but ≤ 1 year vs. CR > 1 year) and prior HSCT versus no prior HSCT. The IWRS data will be compared with the eCRF data for each patient. For cases where there is a difference between the IWRS and eCRF data, the stratification factors as entered in the eCRF will be used in the primary endpoint analysis.

The OS survival curves in each treatment arm will be estimated using Kaplan-Meier methodology and summarized using Kaplan-Meier plots. The treatment effect of idasanutlin + cytarabine versus placebo + cytarabine will be expressed as an HR using a stratified Cox proportional-hazard regression, including a 95% CI. In addition to the OS HR, 6-monthly OS probabilities and the estimated median OS in each arm (if reached) will be presented, including 95% CI.

4.4.2 Secondary Efficacy Endpoints

The fixed-sequence testing procedure will be used (Westfall and Krishen 2001) to adjust for multiple statistical testing of the primary and key secondary efficacy endpoints, thereby controlling the overall type I error rate at a 2-sided significance level of 5%.

The following secondary endpoints will be tested in the *TP53* WT population in the following order provided the null hypothesis of the OS primary endpoint is rejected:

1. CR rate
2. EFS
3. OR (CR, CRp and CRi) rate
4. DOR (duration of CR)
5. Proportion of patients who received HSCT following CR
6. CR rate in mutation defined subgroup: FLT3
7. OS in mutation defined subgroup: FLT3
8. CR rate in mutation defined subgroup: IDH2
9. OS in mutation defined subgroup: IDH2
10. CR rate in mutation defined subgroup: IDH1
11. OS in mutation defined subgroup: IDH1

A given hypothesis in the list above will only be rejected once all previous hypotheses have been rejected at the 2-sided significance level 5%.

Response will be evaluated according to the HMRA criteria (see protocol section 4.5.8) as the best assessment of response over the induction cycle (cycle 1). The frequencies and percentages of patients assigned to each HMRA response category, that is, CR, CRp, CRi, non-responder (partial response [PR], stable disease [SD], resistant disease, relapsed disease, death in aplasia, and death from indeterminate cause), or not available (NA) will be summarized by treatment arm along with the 2-sided 95% CI according to the Pearson-Clopper method.

For the interim futility analysis patients with no response assessments will be considered as NA, unless they directly proceeded to HSCT after Cycle 1. These patients will be considered confirmed CR unless transplantation was clearly performed in relapse.

Proportions of CR as well as ORR (CR, CRp, and CRi) and HSCT will be compared between the two treatment groups using Cochran–Mantel–Haenszel test stratified by the randomization stratification factors. The effect of prognostic factors on these binary endpoints will be assessed in an exploratory analysis using logistic regression.

The time-to-event endpoints, EFS and DOR, will be analyzed using the same statistical methods as described for the primary OS analysis. If very few patients qualify for DOR analysis, only descriptive statistics of DOR will be given.

4.4.3 Sensitivity Analyses:

The following sensitivity analysis will be performed in the *TP53* WT population and all-patient ITT population:

- An unstratified log-rank test for OS
- Discontinuation of assessments or patient lost to follow up considered as an OS event
- An unstratified chi-squared test for CR

A check of the proportional hazards assumption for OS will be performed by plotting the log negative log of the estimated survivor function against log time.

The following sensitivity analysis for OS will be performed in the ITT population:

- A multivariate sensitivity analysis of OS in the ITT population using Cox proportional hazards multiple regression to assess the treatment effect after adjustment for *TP53* mutation status.

To assess the relevance of new treatment for AML against other long-term effects, OS in the *TP53* WT population will be alternatively defined with censoring at date of new treatment and analyzed using the same methods as for the primary OS endpoint

A sensitivity analysis of OS on the subset of patients *TP53* WT (i.e. not including those with retained functionality) will be performed.

To assess the relevance of HSCT against other long-term effects, OS in the *TP53* WT population will be alternatively defined with censoring at date of HSCT and analyzed using the same methods as for the primary OS endpoint.

The confounding effect of HSCT on OS will be assessed using a Cox multivariate regression including major covariates (e.g., prior HSCT, HCT-CI score, leukemia status, type of donor and transplant, engraftment status, conditioning therapy type, site). Only covariates that are significant at 0.15 level in the univariate analysis will then be included in a multivariate analysis, and backward selections at the 5% level will be performed to obtain the model.

Different definitions of EFS will be assessed in sensitivity analyses, in which EFS is defined as

a) time from randomization to the date of treatment failure (failure to achieve CR, set as day 1), relapse from CR, or death from any cause, whichever comes first

b) time from randomization to the date of treatment failure (failure to achieve CR, CRp or CRi, set as day of final response assessment), relapse (from CR, CRp or CRi) or death from any cause, whichever comes first

c) time from randomization to the date of relapse from CR (for subjects who achieved CR) or death from any cause, whichever comes first

d) the above definitions censoring at HSCT.

For time to event analysis the following sensitivity analyses will be performed:

- Patients who report a new treatment post-randomization for follow-up cancer therapy will be considered as having an event (OS and EFS for respective sensitivity analysis) at time of new treatment intake.

- Discontinuation of assessments or patient lost to follow up considered as an event (for OS, and EFS)

4.4.4 Subgroup Analyses

The effects of prognostic factors at baseline, as well as demographic and disease characteristics at baseline, on treatment comparisons for OS (in *TP53* WT ITT population) will be examined and assessed using Cox multivariate regression.

The estimated Kaplan-Meier OS event rates, as well as the HR estimated using an unstratified Cox regression model and 95% CI will be displayed by treatment arm for each level of the subgroup. The subgroup analysis results will be displayed using forest plots (Lewis and Clarke 2001) for stratification factors, the demographic and disease characteristics at baseline, including history of AML.

Descriptive statistics on disease and patient status as well as transplant procedure and engraftment use will be produced to support the analysis of confounding effect of HSCT on OS.

The study is not powered to detect a treatment difference in subgroups. Results from these analyses should be interpreted as exploratory.

4.5 PHARMACODYNAMIC AND BIOMARKER ANALYSES

Exploratory analyses might be performed on bone marrow and blood specimens to identify and/or evaluate the presence of PD biomarker and other biomarkers that might help predict patient subpopulations that are more likely to respond to idasanutlin + cytarabine.

The following PD and biomarker parameters will be presented in patient listings and descriptive summary statistics by treatment arm:

- Mic-1 protein levels in blood samples (analyzed by ELISA)

- *TP53* mutation status
- Rate of CR_{MRD} (MRD response as per ELN guidelines) using multiparameter flow cytometry
- Baseline expression levels of a 4-gene signature will be analyzed by qRT-PCR from patient blood. The predictive or prognostic value of these markers will be analyzed at futility-interim analysis (gene signature) and at the time of primary analysis. These biomarker analyses are exploratory and are not part of the futility interim analysis assessment. However, the iDMC will make a recommendation based on the prognostic and potential value of these markers to inform the decision on how these markers will be further evaluated after the futility-interim analysis; see Section 6.4.4 of the iDMC Charter for further details.
- MDM2 protein expression in blasts may also be analyzed relative to patient outcomes at the end of study.

Details of the biomarker analyses will be described in a separate biomarker analysis plan.

4.6 PHARMACOKINETIC ANALYSES

Details of the PK analyses will be described in the Modeling and Simulation Analysis Plan. Results of this analysis will be reported separately.

4.7 SAFETY ANALYSES

The safety analysis population will be used for all safety analyses (Section 4.1.4).

4.7.1 Exposure of Study Medication

Study treatment exposure (idasanutlin + cytarabine and placebo + cytarabine) will be summarized using descriptive statistics (by treatment arm) for the following measures:

- Number of patients treated overall and by treatment cycle
- Number of days on treatment
- Study medication duration
- Number of treatment cycles started
- Number of doses received and missed doses
- Cumulative dose received
- Dose intensity
- Number and reasons for dose modification and discontinuation

4.7.2 Adverse Events

Coding of adverse events will be done using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA). All adverse events and

laboratory parameters will be assessed according to the National Cancer Institute Common Terminology Criteria for Adverse Event (NCI CTCAE) version 4.03 grading system.

All adverse events will be reported from the day of first study treatment. All adverse events will be displayed by body system (system organ class), NCI CTCAE grade, and relation to study treatment. All serious adverse events, Grade ≥ 3 adverse events, adverse events leading to death, adverse events leading to study treatment modification, adverse events leading to study treatment discontinuations, and adverse events leading to early study withdrawals will be summarized by treatment arm.

Deaths and reason for deaths reported during the study will be summarized by treatment arm. Deaths occurring within 30 and 60 days after randomization will also be summarized.

In the summary tables showing the overall incidence of adverse events, patients who experienced the same event on more than one occasion will be counted only once in the calculation of the event frequency. The adverse event with the most extreme intensity will be used for reporting.

Additional analyses will be performed by treatment arm based on the selected adverse events of special interest, which are defined in the study protocol, as well as based on any adverse event, which might require further investigation. Analyses of the time to onset of these adverse events will be performed. Time to adverse events will be listed. Patients without an adverse event will be censored at last date known to be alive.

In order to account for potential differences in the duration of adverse event reporting between treatment groups, an exploratory analysis of selected adverse events by duration of exposure might be performed based on adverse event rates adjusted for patient-years at risk when enough events are observed.

4.7.3 Laboratory Data

Laboratory data with values outside of the normal ranges will be identified. Laboratory values will be summarized by treatment arm and grade using the NCI-CTCAE version 4.03 grading system. Shift tables will be used to present the number and percentage of patients with values within, above, or below the normal range comparing baseline to subsequent assessments by treatment arm.

Nadir is defined as the lowest value observed in the interval between Cycle 1, Day 1 up to and including predose on Cycle 2, Day 1. If there is no Cycle 2, Day 1 value, and the interval between Cycle 1 Day 1 and Cycle 2 Day 1 is not definable then the latest value in Cycle 1 will be used to define the interval. If no

value is lower than baseline (i.e., Cycle 1, Day 1 cannot be the nadir), then the nadir does not exist, and the patient is excluded from the analysis.

For all patients summary statistics and boxplots for time to nadir (in days) will be produced by treatment arm, as well as Kaplan Meier for time to nadir starting on study drug first administration.

Time to recovery is defined as the time until the value crossed back up and over 100 for platelets and over the thresholds 0.5 or 1 for neutrophils after the interval between Cycle 1, Day 1 and the predose on Cycle 2, Day 1. If there is no Cycle 2, Day 1 value, then the latest value in Cycle 1 will be selected. Patients receiving a transfusion within 24 hours prior to the laboratory assessment selected will be excluded.

For responders only (CR, CRp, CRi), summary statistics and boxplots for time to recovery (in days) for both neutrophils and platelets will be produced by treatment arm, as well as Kaplan Meier for time to recovery starting on study drug first administration.

4.7.4 Vital Signs

Vital signs (absolute values and change from baseline) will be summarized by treatment arm over time without any replacement for missing data.

4.7.5 ECGs

Incidence of clinically significant ECG abnormalities will be reported in patient listings. ECG absolute values and change from baseline will be summarized by treatment arm. ECG QT interval and QT interval with Fridericia's correction will also be summarized.

ECGs will also be analysed in combination with the PK data (area under the concentration–time curve [AUC] and C_{max}) to assess the relationship between study drug plasma concentrations (AUC and C_{max}) and correct QT interval duration.

4.8 PATIENT-REPORTED OUTCOMES

The PRO-evaluable population will include all randomized patients who have a baseline and at least one post-baseline PRO assessment. The PRO-evaluable population will be used for descriptive analyses of visit summary and change from baseline, and responder analyses. All TP53 WT randomized patients will be used for completion analyses and time-to-event (improvement/deterioration) analyses. All PRO analyses will be performed based on the treatment arm assigned at randomization.

4.8.1 EORTC QLQ-C30

The EORTC QLQ-C30 is composed of both multi-item and single-item scales scored 0 to 100. Scoring for the EORTC QLQ-C30 questionnaire will be based on the EORTC guidelines (Fayers et al. 2001). Consistent with the EORTC guidelines, prorated scores will be computed for a multi-item scale if more than 50% of the constituent items have been completed for that scale. For multi-item scales with less than 50% of the items completed, the scale will be considered as missing. Values for missing single-item scales will not be imputed.

Completion analysis will be performed for the overall EORTC QLQ-C30 questionnaire. The number and proportion of patients who completed the EORTC QLQ-C30 questionnaire will be summarized by treatment arm and for each time point. The completion rate will be based on the number of patients expected to complete the questionnaire at a particular time point, that is, those patients who had the opportunity to complete the EORTC QLQ-C30 questionnaire.

Visit summary and change from baseline analyses will be performed for all scales of the EORTC QLQ-C30. Summary statistics (number of patients, mean, standard deviation, median, minimum, maximum) of scores and change from baseline to each time point will be presented by treatment arm. Plots of mean change from baseline will be generated for EORTC QLQ-C30 physical functioning, fatigue, and global health status/QoL scales. In addition, for the subset of patients receiving HSCT, physical functioning and fatigue will be descriptively summarized at their last visit prior to transplant.

For responders' analysis, the number and proportion of patients with a clinically meaningful improvement (as defined below), deterioration, or who remained stable will be summarized by treatment arm, for the EORTC QLQ-C30 physical functioning, fatigue, and global health status/QoL scales. The 95% CI around the proportion will be calculated using the Clopper-Pearson method for each treatment arm. These will be calculated at Cycle 1 Day 28, Cycle 2 Day 28, Cycle 3 Day 28, and at treatment discontinuation

For physical function and global health status/QoL, higher scores are reflective of better HRQoL and a clinically meaningful improvement is defined as at least a 10-point increase. For fatigue, lower scores are reflective of less fatigue and a clinically meaningful improvement is defined as at least a 10-point decrease. Conversely, clinically meaningful deterioration is defined as at least a 10-point decrease and 10-point increase, respectively.

Time to deterioration is defined as the time from randomization to the first documentation of a 10-point decrease in EORTC QLQ-C30 physical functioning and global health status and a 10-point increase in fatigue from baseline. Time to

improvement is defined as the time from randomization to the first documentation of a 10-point increase in EORTC QLQ-C30 physical functioning and global health status and a 10-point decrease in fatigue from baseline.

Time to deterioration and time to improvement will be displayed using Kaplan Meier curves. Patients who do not have an observed deterioration or improvement, respectively, at the time of clinical data cut-off will be censored at the last non-missing assessment date. Patients without a post-baseline assessment will be censored at randomization.

The EQ-5D-5L questionnaire consists of two components: the EQ-5D-5L descriptive system, which has 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with five levels (no problems, slight problems, moderate problems, severe problems, and extreme problems) and a visual analogue scale (EQ VAS) ranging from 0 (worst possible heath state) to 100 (best possible health state), which measures an overall health state.

The EQ-5D-5L health states, which are defined based on the EQ-5D-5L descriptive system, will be converted into a single index value according to the EQ-5D-5L guidelines (van Reene and Janssen 2015). The EQ-5D-5L single index value will be calculated using the United Kingdom preference weights. Descriptive statistics (mean, standard deviation, median, minimum, and maximum) and change from baseline will be presented by treatment arm and at each timepoint.

Data from the EQ-VAS will be used to assess the overall health status. Descriptive statistics (mean, standard deviation, median, minimum, and maximum) and change from baseline at each timepoint will be displayed by treatment arm.

4.9 EXPLORATORY ANALYSES

Primary and secondary efficacy endpoints will be repeated for the all-patient-ITT population using the same methods as described for the TP53 WT population.

Rate of CR_{MRD} will be analysed using same methods as CR.

CR during treatment will be analysed using the same methods as for CR.

Proportion of HSCT *following response* (OR) and HSCT regardless of response will be analysed using the same methods as HSCT following CR.

LFS (Duration of OR) will be analysed using the same methods as DOR.

Depending on the number of patients, exploratory analyses might be performed to assess the possible relationship between idasanutlin consolidation therapy in Cycle 2 and Cycle 3 and the time-to-event measures (OS, EFS, DOR).

Depending on the number of patients who undergo allogeneic HSCT, a landmark OS analysis will be performed with the date of HSCT as baseline using the same methods as for the primary OS endpoint. In addition, a comparison of patients between treatment groups with a similar level of response (i.e., CR, CRi, and CRp) proceeding to transplant or not, might be assessed to describe if there is a difference on their survival outcome.

As a supportive analysis, a competing risk analysis of cause-specific death (with event types "death due to progression of disease" and "death due to adverse event or other reasons") as an exploratory analysis will be performed including the following:

- A visualization of the corresponding cumulative incidence functions
- A cause-specific hazard and Fine-Gray regression models for the competing events

with randomized treatment assignment as the main covariate and potential adjustment for additional potential prognostic variables.

Additional biomarker analyses might be performed as appropriate.

4.10 MISSING DATA

The missing data rules for the primary and secondary efficacy endpoints and for PROs are detailed in Sections 4.4 and 4.8, respectively.

For the reporting of safety data, the following missing data imputation rules will be applied:

- Adverse events with missing grades will be included in the summaries as Grade 3–5 adverse events.
- Adverse events with missing relationship to trial medication will be included in the summaries of treatment related adverse events.
- Adverse events with missing seriousness will be included in all summaries of serious adverse events.

In case of missing day in a time-to-event-date, an imputation is done on the first day of the month or date of randomization, whichever occurs last; otherwise no imputations are performed for the main analysis.

In case of missing assessment of patient primary population (i.e. wild type status), according to FMI an imputation may be done depending on availability of alternative data by Almac, as specified in Section 4.4

Specific sensitivity analyses are going to be performed on missing data as specified in Section 4.4.3.

4.11 INTERIM ANALYSES

The iDMC will perform regular safety review, assess the lack of efficacy (futility) for this study, and perform the interim efficacy analysis (Section 3.2) on testing OS. iDMC recommendations to stop the study due to safety, or because of substantial evidence of lack of efficacy of the study drug, and for early unblinding for OS-efficacy, are based on the pre-specified interim analysis methodology as defined in the iDMC charter. The iDMC will provide their recommendation to the Sponsor. The Sponsor will make the final decision regarding the early termination of the study for safety or futility or early unblinding for efficacy.

All summaries and analyses will be prepared by an Independent Data Coordinating Center (IDCC) and presented by treatment arm for the iDMC's review as described in the iDMC charter. Members of the IDCC and the iDMC will be external to the Sponsor and the study management team. The IDCC and iDMC members will comply with the iDMC charter that outlines their roles and responsibilities.

4.11.1 Futility Interim Analysis

The non-binding futility interim analysis will be based on confirmed CR, which is defined as CR or CRp with a duration of least 28 days post initial response assessment; EFS; and safety. Note that EFS definition at time of futility analysis (treatment failure defined as failure to achieve CR, CRp or CRi) was different to that now proposed at final analysis, due to comments by regulatory agencies.

The futility interim analysis will be performed on the interim analysis population (defined in Section 4.1.1.1). The iDMC may recommend stopping the study for futility if, in the sample of patients who are wild type for *TP53 (as assessed by the laboratory:* Almac Diagnostic Services*)*, either of the following is observed:

- The observed odds ratio for confirmed CR, which compares idasanutlin + cytarabine to placebo + cytarabine, is < 2.0.

  or

- The observed odds ratio for confirmed CR, which compares idasanutlin + cytarabine to placebo + cytarabine, is < 2.5, and the HR for EFS is > 1.

The iDMC recommendation for this trial to not be considered futile is based on the observed odds ratio CR ≥ 2.5, which sets a high-bar target for more than doubling the CR, assuming 16% CR in the placebo + cytarabine arm versus 32.3% CR in the idasanutlin + cytarabine arm. If the observed odds ratio falls between 2 and 2.5, then EFS is included as an event time-based measure of activity for the lower gate of odds ratio, and its HR should be ≤ 1 to continue. Durable CR, as described above, was chosen instead of a survival endpoint because OS carries the risk of incorrect decision-making being confounded by treatment-

related mortality at this early time of the study. EFS (as defined in Section 4.4.2) was added to support the association between CR and OS.

To compute stopping probabilities for the futility interim analysis, the following additional assumptions are made within *TP53* WT patients and the simulation model is extended accordingly:

- Median OS for non-responders in cytarabine arm is 5.1 months
- Median OS for responders, but short-term survivors in cytarabine arm is 7.5 months
- EFS follows an exponential distribution.
- Median EFS times for non-responders and CR short-term responders is assumed to be shorter by a factor 2.5 compared to OS in these same subpopulations.
- The correlation between uncensored EFS and OS times is 0.5.
- HR for a comparison of idasanutlin + cytarabine versus cytarabine + placebo in both, non-responders and short-term responders, is 0.8

Approximately 63 EFS events are expected to occur at the interim analysis under the alternative hypothesis of an increase in median OS from 6 to 9 months. The probability of early stopping due to futility is approximately 89.9% if the null hypothesis of equal OS survival functions is true and is approximately 31.9% if the alternative assumption is true.

4.11.2 Safety Interim Analysis

The iDMC might also recommend stopping the study for safety at the futility interim analysis if any of the following criteria are met in all patients, regardless of *TP53* mutational status (note: early death is defined as any death within the first 30 days after randomization):

- The proportion of gastrointestinal toxicity (nausea, vomiting, diarrhea) events in the idasanutlin + cytarabine treatment arm is > 40% of events that are Grade 3 or > 15% of events that are Grade 4

  or

- The proportion of early deaths in the idasanutlin + cytarabine treatment arm is ≥ 10 percentage points greater than in the placebo + cytarabine treatment arm.

  or

- > 20% of early deaths overall in the idasanutlin + cytarabine treatment arm

4.11.3 Efficacy Interim Analysis

At the time of the efficacy interim analysis (on OS) that will be conducted when 80% of the events have occurred (i.e., approximately 220 events), it is

anticipated that all patients have been enrolled. OS will be tested at the significance level determined using the O´Brien-Fleming alpha-spending function so the overall Type I error rate will be maintained at the 0.05 level. With 80% information, the alpha spending is 0.024.

The iDMC may recommend unblinding the study for efficacy if there is a significant difference in OS in favor of the experimental arm (alpha level 0.024 or as assessed based on actual number of OS events).

Accompanying this interim efficacy analysis, a non-binding additional futility assessment will be performed and the iDMC may recommend stopping the study for futility if the hazard ratio on overall survival is greater than or equal to one.

Further details about the interim analyses will be described in the iDMC Charter.

5. REFERENCES

Dohner H, Estey E, Grimwade D et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129:424-447.

Fayers PM, Aaronson NK, Bjordal K, et al, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 scoring manual. 3rd ed. Brussels: European Organisation for Research of Cancer, 2001.

Lewis S, Clarke M. Forest Plots: trying to see the wood and the trees. BMJ 2001;322:1479–80.

Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2018;131:1275-1291.

Van Reene M, Janssen B, editors. EQ-5D-5L User Guide: Basic informtaion on how to use the EQ-5D-5L instrument. Version 2.1. The Netherlands: EuroQol Research Foundation, April 2015.

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

Appendix 1

PROTOCOL SYNOPSIS

TITLE:	**A MULTICENTER, DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED, PHASE III STUDY OF IDASANUTLIN, AN MDM2 ANTAGONIST, WITH CYTARABINE VERSUS CYTARABINE PLUS PLACEBO IN PATIENTS WITH RELAPSED OR REFRACTORY ACUTE MYELOID LEUKEMIA (AML)**
PROTOCOL NUMBER:	WO29519
VERSION NUMBER:	6
EUDRACT NUMBER:	2014-003065-15
TEST PRODUCT:	Idasanutlin (RO5503781)
PHASE:	Phase III
INDICATION:	Relapsed/Refractory Acute Myeloid Leukemia
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives

Primary Objective

The primary objective for this study is as follows:

Within the *TP53* wild type (WT) population

- To compare overall survival (OS) in patients with relapsed or refractory acute myeloid leukemia (AML) who have been randomized to idasanutlin in combination with cytarabine versus those who have been randomized to cytarabine and placebo

Secondary Objective

The secondary objectives for this study are as follows:

Within the *TP53* WT population

- To compare the proportions of complete remission (CR) between treatment arms
- To compare event-free survival (EFS) between treatment arms
- *To compare ORR (defined as CR), complete remission with incomplete platelet count recovery [CRp], and complete remission with incomplete blood count recovery [CRi]) between treatment arms*
- *To compare duration of remission following CR (DOR) between treatment arms*
- To compare the proportions of allogeneic hematopoietic stem cell transplant (HSCT) following *CR* between treatment arms
- *To assess OS and CR in clinically actionable mutation-defined AML subpopulations, including FLT3, IDH1, and IDH2*
- *To assess the safety of idasanutlin plus cytarabine as compared with cytarabine and placebo*

- *To compare the differences in disease and treatment-related symptoms and health-related quality of life between treatment arms*

Within the all-patient population

- To assess the safety of idasanutlin plus cytarabine as compared with cytarabine and placebo
- To characterize the pharmacokinetics of both idasanutlin and cytarabine
- To compare the differences in disease and treatment related symptoms and health-related quality of life between treatment arms

Exploratory Objective

The exploratory objectives for this study are as follows:

Within the all-patient population

- *To evaluate primary and secondary efficacy endpoints, including OS, CR, proportion of HSCT, EFS, and DOR in the all-patient population and OS and CR in the clinically actionable mutation-defined AML subpopulations (including FLT3, IDH1, and IDH2)*

Within the all-patient and TP53 WT populations

- *To evaluate efficacy endpoints, including proportion of HSCT, CR, ORR, and DOR on the basis of response assessed over complete treatment period*
- *To evaluate LFS (leukemia-free survival, or duration of OR)*
- To explore minimal*/measurable* residual disease (MRD) in the bone marrow after treatment with idasanutlin plus cytarabine compared with cytarabine and placebo
- To evaluate candidate response biomarkers (gene *expression signatures*) and murine double minute 2 (MDM2) protein expression in blast cells
- *To assess prognostic and predictive effect of disease-associated mutations*

Study Design

Description of Study

This is a Phase III multicenter, double-blind, randomized, placebo-controlled study of idasanutlin in combination with cytarabine compared with cytarabine and placebo.

A total of 440 patients with AML who have relapsed following, or are refractory to cytarabine–containing standard induction chemotherapy after at least one and no more than two prior cytarabine–containing induction chemotherapy regimen(s) are planned to be enrolled. Relapsed patients are defined as patients with first or second relapse; first relapsed patients who are young and had a good response to initial therapy (i.e. age < 60 years with first CR achieved [CR1] duration > 1 year) are excluded. Refractory patients are defined as patients with persistent leukemia after one or two induction cycles, or patients with CR1 duration of < 90 days. Patients may have received prior HSCT in remission. Note that patients with prior allogenic HSCT within 90 days prior to randomization will not be eligible for this study. The *TP53* WT population will consist of patients with WT *TP53*, established centrally.

Re-screening is allowed under the conditions listed in the protocol.

Patients will be randomly assigned to each treatment arm. The arms will be stratified according to the following factors:

- Age (< 60 years versus ≥ 60 years)
- Cytogenetic and molecular risk according to *the 2010* European LeukemiaNet (ELN) standardized reporting system at initial diagnosis (favorable/intermediate versus adverse). Cytogenetic information allowing stratification in those two groups must be available.
- Response to initial anti-leukemic therapy (refractory versus CR ≥ 90 days but ≤ 1 year; versus CR > 1 year)
- Prior HSCT versus no prior HSCT

Patients will receive induction treatment with idasanutlin/placebo 300 mg twice daily (BID) and cytarabine 1 g/m^2 for 5 days followed by 23 days of rest (Cycle 1). For patients who achieve clinical response after induction and for whom HSCT for consolidation is not an option, it is strongly recommended to continue study medication with a maximum of two additional cycles of consolidation (Cycles 2 and 3) unless clinically contraindicated

Responding patients, including those who proceed to HSCT, will be followed for EFS and *DOR*. All patients irrespective of response to treatment will be followed for OS until the end of the study.

An interim analysis for futility based on CR, EFS, and safety is planned after 120 *TP53* WT patients are enrolled, have received at least one cycle, and confirmatory response assessment is available, to allow early stopping of the study in the event of inadequate CR, EFS, or safety concerns. The *futility* interim analysis will be performed by an independent Data Monitoring Committee (iDMC). Enrollment of patients will continue while *this* interim analysis takes place. If the Sponsor decides to stop the study based on iDMC recommendation following the *futility* interim analysis, enrollment will be discontinued.

An efficacy interim analysis on OS is planned to be conducted by iDMC to allow early testing of OS in the event of relevant efficacy. The efficacy interim analysis will be conducted when 80% of the OS events have occurred (i.e., approximately 220 events). At this time, it is anticipated that all patients will have been enrolled.

Accompanying this interim efficacy analysis, a non-binding additional futility assessment will be performed based on a hazard ratio on overall survival ≥1.

In the case of not reaching significance for OS at interim analysis, a primary analysis of OS will be performed (i.e., when 275 events have occurred).

Number of Patients

Overall, 440 patients with relapsed or refractory AML are expected to be enrolled in this study, over approximately *41* months.

Target Population

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Age ≥ 18 years
- Documented/confirmed first or second refractory or relapsed AML using World Health Organization classification, except acute promyelocytic leukemia. Please note that first relapsed AML patients with CR1 duration of > 1 year AND age < 60 years are excluded.
- No more than 2 prior induction regimens (excl. prior HSCT) in their first line treatment, and one must have included cytarabine with an anthracycline (or anthracenedione).
- Eastern Cooperative Oncology Group performance status of 0–2
- Adequate hepatic function assessed by the following:

  Serum total bilirubin ≤ 1.5 × institutional upper limit of normal (ULN), unless resulting from hemolysis, Gilbert's syndrome, or liver infiltration with leukemia

  AST/ALT ≤ 3 × institutional ULN (or ≤ 5 × upper limit of institutional laboratory reference range if liver infiltration with leukemia)
- Adequate renal function assessed by serum creatinine within reference laboratory ranges OR creatinine clearance (by Cockcroft Gault formula) ≥ 50 mL/min
- WBC count at randomization of ≤ 50,000/mm^3

  Note: When treatment is not started immediately upon randomization, the WBC count at the start of induction therapy (Cycle 1) must remain at ≤ 50,000/mm^3. The use of hydroxyurea (HU) or leukapheresis to meet eligibility is allowed. HU or leukapheresis must be discontinued at least 24 hours prior to the initiation of study medication.
- For women of childbearing potential: agreement to remain abstinent (refrain from heterosexual intercourse) or use two adequate methods of contraception, including at least one method with a failure rate of < 1% per year, during the treatment period and for up to 6 months after the last dose of study drug

A woman is considered to be of childbearing potential if she has not reached a postmenopausal state (≥ 12 months of amenorrhea with no identified cause other than menopause), and has not undergone surgical sterilization (removal of ovaries and/or uterus).

Examples of contraceptive methods with a failure rate of < 1% per year include bilateral tubal ligation, male sterilization, and established, proper use of hormonal contraceptives that inhibit ovulation, hormone-releasing intrauterine devices, and copper intrauterine devices.

The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception. Barrier methods must always be supplemented with the use of a spermicide.

- For men unless permanently sterile by bilateral orchidectomy: agreement to remain abstinent (refrain from heterosexual intercourse) or use contraceptive measures and agreement to refrain from donating sperm, as defined below:

  With female partners of childbearing potential or pregnant female partners, men must remain abstinent or use a condom during the treatment period and for up to 6 months after the last dose of study drug. Men must refrain from donating sperm during this same period.

  The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient.
- Ability to understand and willingness to sign a written informed consent form and comply with all study requirements including completion of patient-reported outcome measures.

Exclusion Criteria

Patients who meet any of the following criteria will be excluded from study entry:

- First relapsed patients aged < 60 years with a CR1 duration of > 1 year
- Patients with prior documented antecedent hematological disorder including the following: myelodysplastic syndrome, myeloproliferative disease (i.e., chronic myelomonocytic leukemia, polycythemia vera, primary myelofibrosis, and essential thrombocythemia), and aplastic anemia
- AML secondary to any prior chemotherapy unrelated to leukemia
- Patients who are either refractory to or have relapsed within 90 days of receiving a regimen containing a cumulative dose of ≥ 18 g/m^2 cytarabine
- Patients who have received allogeneic HSCT within 90 days prior to randomization. HSCT should have been performed in remission and not used for salvage (patients who have received autologous HSCT as consolidation in CR1 are eligible).
- Patients who have received immunosuppressive therapy for graft-versus-host disease or for engraftment syndrome after autologous stem cell transplantation within 2 weeks prior to randomization
- Prior treatment with an MDM2 antagonist
- Patients with clinically relevant QTc prolongation (QT interval corrected using Fridericia's formula [QTcF] > 480 ms), a family history of long QT syndrome, or who are currently receiving treatment with medications that are known to prolong the QT interval

  Medications that are known to prolong the QT interval must be discontinued 7 days (or 5 half-lives, whichever is shorter) prior to initiating study medication until 5 days after the final administration of study medication.

- Patients receiving any other investigational or commercial agents or therapies administered with the intention to treat their malignancy within 30 days (or 5 half-lives) from first receipt of study drug

  Note: The exception is HU or leukapheresis in patients who need to continue this therapy to maintain a WBC count $\leq$ 50,000/mm^3. HU or leukapheresis must be discontinued at least 24 hours prior to the initiation of study medication.
- Patients with acute toxicities from any prior anti-leukemia therapy which have not resolved to Grade $\leq$ 2 per National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE), Version 4.03
- Patients with a history of other malignancy within 5 years prior to screening, except for malignancy that has been in remission without treatment for at least 2 years prior to randomization
- Patients unable to temporarily interrupt treatment with moderate to strong CYP2C8 inducers and inhibitors (including gemfibrozil, which is also an inhibitor of UGT1A3), CYP2C8 or OATP1B1/3 substrates, or strong CYP3A4 inducers during the treatment phase. These agents must be discontinued 7−14 days prior to the start of study medication.
- Patients unable to temporarily interrupt treatment with oral or parenteral anticoagulants/anti-platelet agents (e.g., warfarin, chronic daily treatment with aspirin [$>$ 325 mg/day], clopidogrel, dabigatran, apixaban, rivaroxaban) during the treatment phase. These agents must be discontinued 7 days (or 5 half-lives) prior to the start of study medication.

  Note: Treatment with or switch to low molecular weight heparin (LMWH) or unfractionated heparin (UFH) is allowed, according to local practice. However, platelet levels need to be closely monitored in these patients (see protocol).
- Patients with a history of systemic hypersensitivity reactions $\geq$ Grade 2 attributed to cytarabine or components of the formulated product
- Patients who have any severe and/or uncontrolled medical conditions or other conditions that could affect their participation in the study, impair the ability of the investigator to evaluate the patient, or impair the patient's ability to complete the study such as the following:

  Unstable angina, symptomatic or otherwise uncontrolled arrhythmia (does not include stable, lone atrial fibrillation), uncontrolled hypertension, symptomatic congestive heart failure (New York Heart Association III, IV), myocardial infarction $\leq$ 6 months prior to first study medication, and cerebrovascular accidents $\leq$ 6 months before study medication start

  Unstable seizure disorders

  Nonmalignant medical illnesses that are uncontrolled or whose control may be jeopardized by this study medication, such as hereditary coagulation disorders or insulin-dependent diabetes mellitus not optimally controlled with medical management (e.g., presence of ketoacidosis) or active gastrointestinal (GI) conditions (e.g., Grade $\geq$ 2 graft−versus-host disease) and uncontrolled $inflammatory$ bowel disease (i.e., Crohn's disease, ulcerative colitis, diverticulosis-associated colitis, and Behçet's disease).
- Infection considered by the investigator to be clinically uncontrolled or of unacceptable risk to the patient upon the induction of neutropenia, that is, patients who are or should be on antimicrobial agents for the treatment of active infection such as the following:

  Fungal infection with visceral involvement, other than mucosal candidiasis, with $<$ 2 weeks of appropriate systemic antifungal therapy

  Active bacterial infection and/or bacterial infection with positive cultures in the 7 days prior to dosing

  Patients who have received $<$ 5 days of appropriate therapeutic antibiotic therapy for an identified infection

History of symptomatic *Clostridium difficile* infection that required treatment within 1 month prior to dosing. Upon clinical response to *C. difficile* treatment, the stool consistency and frequency must have returned to normal.

In all cases, the patient should be afebrile (exception of AML-related fever) and hemodynamically stable for at least 72 hours at the time of study medication initiation.

- Patients with a history of active or chronic infectious hepatitis unless serology demonstrates clearance of infection

  Patients with occult or prior hepatitis B virus (HBV) infection (defined as negative hepatitis B surface antigen and positive total hepatitis B core antibody) may be included if HBV DNA is undetectable, provided that they are willing to undergo monthly DNA testing. Patients who have protective titers of hepatitis B surface antibody after vaccination or prior but cured hepatitis B are eligible. Patients positive for hepatitis C virus antibody are eligible provided polymerase chain reaction (PCR) is negative for HCV RNA.
- Patients who have a history of clinically significant liver cirrhosis (e.g. Child-Pugh class B and C).
- Patients with electrolyte abnormalities such as hypokalemia, hyperkalemia, hypocalcemia, hypercalcemia, hypomagnesemia, and hypermagnesemia of Grade $>$ 1 per NCI CTCAE v4.03. Treatment for correction of above electrolyte imbalances is permitted during screening to meet eligibility.
- Patients with extramedullary AML with no evidence of systemic involvement
- Patients with active CNS leukemia
- Pregnant or breastfeeding patients
- HIV-positive patients
- Patients who might refuse to receive blood products and/or have a hypersensitivity to blood products

End of Study

The end of this study is defined as the date when the last patient, last visit (LPLV) occurs. LPLV is expected to occur 2 years after the last patient is enrolled or after all patients have died, whichever occurs first.

Length of Study

The total length of the study, from screening of the first patient to the end of the study, is expected to be approximately 5.5 years.

Investigational Medicinal Products

Test Product (Investigational Drug)

Idasanutlin/placebo will be administered at 300 mg BID orally, without regard to meals. Water can be given ad libitum.

In case the patient proceeds to consolidation therapy, only 50% of idasanutlin should be given (300 mg once daily [QD] in the morning).

Cytarabine will be administered at a dose of 1 g/m^2 QD as a 1–3 hour intravenous infusion. In case the patient proceeds to consolidation therapy the dose may need to be reduced to 50% (0.5 g/m^2 QD).

Statistical Methods

Efficacy Analysis

Patients will be analyzed according to the treatment arm to which they were randomized. *The TP53 WT population is the efficacy population and refers to all randomized TP53 WT patients as identified by a central laboratory test.*

Description of efficacy analysis in this section is valid for both interim and primary analysis.

Primary Efficacy Endpoint

The primary efficacy endpoint, OS, is defined as the time from randomization to death due to any cause. OS for patients who have not died at the time of the analysis will be censored at the date last known alive.

The primary OS analysis of the study will assess the null hypothesis of equality of OS functions in the idasanutlin in combination with cytarabine (MDM2-chemo) arm versus the cytarabine given with a placebo (chemo) arm in the *TP53* WT population as follows:

H0: $OS_{MDM2\text{-}chemo} = OS_{chemo}$ versus H1: $OS_{MDM2\text{-}chemo} \neq OS_{chemo}$

A formal treatment comparison will be made using a two-sided stratified log-rank test at a significance level *defined using the O´Brien-Fleming alpha-spending function with overall Type I error rate at 0.05 corresponding to available information.* Stratification factors that will be used are the same as for randomization, i.e. age (< 60 versus ≥ 60 years), cytogenic and molecular risk (favorable/intermediate versus adverse), prior response to initial anti-leukemic therapy (refractory versus CR ≥ 3 months but ≤ 1 year versus CR > 1 year) and prior HSCT versus no prior HSCT.

Survival curves in each treatment arm will be estimated using Kaplan−Meier estimates. The Kaplan–Meier estimates will provide a visual description of the survival curves and the difference across treatment arms. The treatment effect will be quantified via a hazard ratio, computed from a stratified Cox proportional-hazards regression, including a 95% CI. To further describe and quantify OS, estimated median OS per arm and 1-year and 2-year survival probabilities, all including 95% CI, will be given. The effect of prognostic factors on OS will be assessed in an exploratory analysis using Cox multivariate regression.

Secondary Efficacy Endpoint

The following secondary endpoints will be tested for the TP53 WT population, as described in the Statistical Analysis Plan:

- *CR proportion*
- EFS
- ORR (CR, CRp, and CRi)
- *DOR (duration of remission following CR)*
- Proportion of HSCT *following CR*
- *CR proportion and OS in clinically actionable mutation-defined subpopulation (FLT3, IDH1, and IDH2)*

Patients with no response assessments (for any reason) will be considered non-CR.

Difference in proportions of CR will be assessed between the two treatment arms using Cochran–Mantel–Haenszel test stratified by randomization stratification factors. In addition, proportions and 95% CI will be reported for each treatment arm. The effect of prognostic factors on CR will be assessed in an exploratory analysis using logistic regression.

EFS is defined for all patients and measured from the date of randomization. It is measured until treatment failure, relapse from CR, or death from any cause, whichever occurs first. For patients with none of these events before *time of analysis,* EFS is censored at the date of the patient's last response assessment.

DOR is defined for patients achieving complete remission and is the time from clinical remission until relapse or death from any cause, whichever occurs first. For patients with none of these events before time of analysis, DOR is censored at the date of the patient's last response assessment.

EFS and *DOR* will in general be analyzed using the same statistical methods as those described for OS. If only very few patients qualify for *DOR* analysis, only descriptive statistics will be given.

ORR (CR, CRp, and CRi) and proportion of HSCT will be compared between the two treatment arms using the same statistical methods as those described for CR.

Sensitivity Analyses

The following sensitivity analyses for OS will be performed in the *TP53* WT population:

- An unstratified log-rank test.
- To assess the relevance of HSCT against other long-term effects, OS will be alternatively defined with censoring at date of HSCT and analyzed using the same methods as for the primary endpoint.
- *Discontinuation of assessments or patient lost to follow up considered as an event*

Safety Analyses

All safety analyses will be based on *both the TP53 WT population (defined here as TP53 WT patients who have received any study medication at least once) and* the complete safety analysis population (defined as all patients who have received any study medication at least once), and patients will be analyzed according to the treatment received (patients receiving idasanutlin at least once will be analyzed in the idasanutlin arm). Safety analyses will include, but not be limited to, incidence rates for adverse events including mortality, adverse event severity, seriousness, and adverse events leading to discontinuation. In addition, abnormalities of clinical laboratory tests and vital signs assessed during the study treatment period and post-treatment follow-up will be assessed. Exposure to study medication will be summarized by total duration of study medication, number of cycles started and cumulative dose using descriptive statistics.

This trial is designed to allow for early termination or a modification of the protocol for safety concerns or lack of efficacy, based on the advice of an iDMC. The iDMC will be incorporated into the study to review safety data on a regular basis, including adverse events of special interest. Both the Sponsor and the iDMC can request ad hoc iDMC meetings if potential safety concerns arise. Following each meeting, the iDMC will recommend to the Sponsor whether the study should continue according to the protocol or may suggest changes to the protocol based on the outcome of the data review. In exceptional cases, the iDMC may recommend stopping the study or closing a treatment arm as a result of safety reasons. The iDMC will also perform a safety review at the preplanned interim analyses for futility and efficacy.

Pharmacodynamic and Biomarker Analyses

The following pharmacodynamic parameters will be presented by listings and descriptive summary statistics.

- Blood samples analyzed for macrophage inhibitory cytokine-1
- Analysis of *TP53, FLT3, IDH1, and IDH2* mutation status
- *MDM2 protein expression level in AML blasts*
- *A 4-gene signature (including MDM2 gene expression)*
- *MRD*

Pharmacokinetic Analyses

Key pharmacokinetic (PK) parameter values (apparent clearance and apparent volume of distribution) of idasanutlin and cytarabine (total clearance and volume of distribution) will be estimated using a population pharmacokinetics (popPK) approach. The influence of covariates such as gender, race/ethnicity, weight, hematological parameters at baseline, renal/hepatic impairment, and degree of underlying disease will be investigated. Other PK parameters such as maximum concentration observed (C_{max}), steady-state concentration at the end of a dosing interval (i.e., just prior to next drug administration) (C_{trough}), area under the concentration–time curve during one dosing interval, area under the concentration-time curve during a 24-hour dosing interval (AUC_{0-24h}), and half-life will be derived from the individual post hoc predictions. Details of the population analysis will be described in the Modeling and Simulation Analysis Plan. Results of this analysis will be reported separately.

If appropriate, an exploratory PK/pharmacodynamic analysis may be performed post hoc. The primary focus will be exploration of the relationship between measures of exposure to idasanutlin in combination with cytarabine (C_{max}, C_{trough}, and AUC_{0-24h}) and ECG, drug-related adverse effects as well as clinical efficacy parameters.

Determination of Sample Size

A mechanistic simulation model *was* used to determine the sample size in this event-driven study, based on the following global assumptions:

- Final analysis for OS in patients with *TP53* WT disease based on two-sided log-rank test at 0.05 level of significance
- 85% power to detect an OS hazard ratio for idasanutlin + cytarabine versus cytarabine + placebo of 0.67 in patients with *TP53* WT disease, corresponding to an improvement in median OS from 6 to 9 months (50%)
- Proportion of long term survivors of 8.0% in the cytarabine + placebo arm and 16.1% in the idasanutlin + cytarabine arm

To compute the necessary number of events, we *simulated* OS times based on the following assumptions:

- All simulated OS times for patients not considered long-term survivors are exponentially distributed.
- Probability of being a complete responder in the cytarabine + placebo arm is 0.16
- Probability of being a complete responder in the idasanutlin + cytarabine arm is 0.323, implying an odds ratio for CR comparing the idasanutlin + cytarabine versus the cytarabine + placebo arm of 2.5
- Probability for a complete responder to be a long-term survivor is 0.5 in either arm
- An annual dropout rate of 5% (every effort will be made to contact patients for survival information in case of study withdrawal or loss to follow-up)

To have the targeted 85% power, 275 events *in patients with TP53 WT disease* are required. The minimum detectable hazard ratio in a 2:1 randomized trial corresponding to 275 events and a significance level of 0.05 amounts to 0.78, corresponding to a minimal detectable median improvement from 6 to 7.7 months assuming exponentiality.

All patients*, regardless of TP53 mutation status,* will be randomized to this *study. Assuming 85% of patients will have TP53 WT disease and 15% will have TP53 mutant disease, approximately* 440 patients will be enrolled over approximately 29 months, corresponding to an estimated number of 374 patients *with TP53 WT disease.*

In Version 6 of the protocol, an interim analysis for efficacy on OS of TP53 WT patients was added. The sample size assumptions above remain unchanged other than the recruitment time (now estimated at 41 months). The interim analysis will occur at an information fraction of 80%, providing a maximum power of 83% for the final OS analysis.

Interim *Analyses*

Futility

A non-binding interim analysis for safety and futility will be performed by an iDMC after 120 patients with *TP53* WT have been enrolled and assessed for response. For the purposes of the *futility* interim analysis, CR is defined as confirmed CR. Sponsor personnel will not have access to by-arm efficacy and safety summaries prior to the formal reporting of study results. The iDMC may recommend stopping the study for futility if:

- the observed odds ratio for CR in the cytarabine and idasanutlin arm versus the cytarabine and placebo arm in the population of patients WT for TP53, is < 2.0,
- or the observed odds ratio for CR in the cytarabine and idasanutlin arm versus the cytarabine and placebo arm in the population of patients WT for TP53, is < 2.5 and the hazard ratio for EFS > 1.

To compute stopping probabilities for the *futility* interim analysis, the following additional assumptions are made within *TP53* WT patients and the simulation model is extended accordingly:

- Median OS for non-responders in cytarabine arm is 5.1 months
- Median OS for responders, but short-term survivors in cytarabine arm is 7.5 months
- EFS follows an exponential distribution

- Median EFS times for non-responders and CR short-term responders is assumed to be shorter by a factor 2.5 compared to OS in these same subpopulations
- The correlation between uncensored EFS and OS times is 0.5
- Hazard ratio for a comparison of idasanutlin + cytarabine versus cytarabine+ placebo in both non-responders and short-term responders is 0.8

Using these assumptions, we expect 63 EFS events at the *futility* interim analysis under the alternative hypothesis. The probability of early stopping due to futility is 89.9% if the null hypothesis of equal OS survival functions is true and 31.9% if the alternative assumption of an increase in median OS from 6 to 9 months is true.

The analysis in both the *TP53* WT population and in the overall population will be provided to the iDMC.

The iDMC may recommend stopping the study for safety at the *futility* interim analysis if any of the following criteria are met (note: early death is defined as any death within the first 30 days after randomization):

- The proportion of GI toxicity (nausea, vomiting, diarrhea) events in the experimental arm (idasanutlin + cytarabine): Grade 3 $>$ 40% or Grade 4 $>$ 15%
- The proportion of early deaths in the experimental arm (idasanutlin + cytarabine) is $\geq$ 10 percentage points greater than in the control arm (cytarabine + placebo)
- $>$ 20% of early deaths overall in the treatment arm

More details on the interim analysis are provided in the iDMC charter.

Efficacy

An interim analysis for efficacy on OS is planned to be conducted by iDMC. The efficacy interim analysis will be conducted when 80% of the OS events have occurred (i.e., approximately 220 events). At this time, it is anticipated that all patients will have been enrolled.

For the interim efficacy analysis of OS, the significance level will be determined using the O'Brien-Fleming alpha-spending function with overall type I error rate at 0.05 level. At the time of the interim efficacy analysis, it is expected that 80% of the OS events will have occurred, corresponding to an alpha spending of 0.025 for the interim and 0.043 for the final OS analysis leading to a power of 83% for the final analysis.

The iDMC will test OS for efficacy and check if there is a significant difference (alpha level 0.025 or as assessed based on actual number of OS events) in OS in favor of the experimental arm.

Accompanying this interim efficacy analysis, a non-binding additional futility assessment will be performed and the iDMC may recommend stopping the study for futility if the hazard ratio on overall survival is greater than or equal to one.

Further details of the interim analyses will be described in the iDMC Charter and Statistical Analysis Plan.

Appendix 2

Calling Rules to Define *TP53* WT Intent-to-treat Primary Population

Variant Type	**Description**	**Assay* Reporting Result**
No variant	Absence of any change in sequence from reference sequence	*TP53*WT
Synonymous substitution	Silent mutations where the amino acid sequence is unchanged	*TP53*WT
Known germline variants	Inclusion in dbSNP version 135 filtered by inclusion in 1000 Genomes phase 1 or HapMap with some custom filtering	*TP53*WT
Missense - Fully Functional	Missense mutation that are categorized as "fully functional" in IARC database under transactivation class; http://p53.iarc.fr/TP53GeneVariations.aspx	*TP53* mutant with retained functionality

*TP53 mutation status was initially assessed using a single target NGS assay (ILLUMINA Truesight platform, customized for TP53) by Almac Diagnostic Services. Following the interim futility analysis, the assay was switched to the F1CDx panel by Foundation Medicine Inc. (FMI), to afford a regulatory path for companion diagnostic (CDx) development and to test other clinically actionable mutations. All available samples previously tested by ILLUMINA assay have been retrospectively tested with the F1CDX assay. Going forward, concordance of results with ILLUMINA is being established for all available samples to support the CDx assay validation.

The primary population will be defined from FMI data. The Almac data will only be used if FMI data are not available; In all cases the above described calling rules will be applied. =