TOLERO PHARMACEUTICALS
Sumitomo Dainippon Pharma Global Oncology

STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 2, Randomized, Biomarker-Driven, Clinical Study in Patients with Relapsed or Refractory Acute Myeloid Leukemia (AML) with an Exploratory Arm in Patients with Newly Diagnosed High-Risk AML and Exploratory Arms with Varying Levels of MCL-1 Dependence
Protocol Number:	TPI-ALV-201
IND Number:	057729
Study Drug:	Alvocidib (test agent; formerly flavopiridol) in combination with Ara-C (cytarabine) and mitoxantrone
Phase of Development:	Phase 2
Sponsor:	Tolero Pharmaceuticals, Inc. [redacted]
Date of SAP:	13-JUN-2018 (Version 1.0)

Table of Contents

List of Tables

List of Appendices

1. Synopsis

Protocol Title:

A Phase 2, Randomized, Biomarker-Driven, Clinical Study in Patients with Relapsed or Refractory Acute Myeloid Leukemia (AML) with an Exploratory Arm in Patients with Newly Diagnosed High-Risk AML and Exploratory Arms with Varying Levels of MCL-1 Dependence

Study Design:

There is a primary study component and four exploratory study arms

- Primary Study: single-arm, Simon two-stage lead in to confirm efficacy assumptions followed by a randomized, two-arm parallel design
- Exploratory Study Arms: single-arm observational

Study Populations:

- Primary Study: Relapsed and refractory AML patients with MCL-1 dependence ≥40%
- Exploratory Study Arms:
  - Patients with newly diagnosed, high-risk AML and MCL-1 dependence ≥40%
  - Relapsed and refractory AML patients with MCL-1 dependence between 0% and <15%
  - Relapsed and refractory AML patients with MCL-1 dependence between 15% and <30%
  - Relapsed and refractory AML patients with MCL-1 dependence between 30% and <40%

Treatment Arms:

- ACM (alvocidib, Ara-C [cytarabine] and mitoxantrone)
- CM (Ara-C [cytarabine] and mitoxantrone)

Biomarker:

- MCL-1 dependence measured by mitochondrial sensitivity to NOXA BH3 peptides (hereafter referred to as MCL-1 dependence)

Sample Size:

- Primary Study, Stage 1: up to 23 patients
- Primary Study, Stage 2: 106 patients (53 patients per treatment arm)
- Exploratory Study Arms: 20 patients in each of four exploratory patient populations

Randomization:

- Primary Study, Stage 1: Not randomized, all patients receive ACM

- Primary Study, Stage 2: Randomized 1:1 to ACM or CM stratified by response to frontline AML therapy
- Exploratory Study Arms: Not randomized, all patients receive ACM

Analysis Populations:

Intent-to-treat (primary for efficacy), safety, response-available, per-protocol, crossover

Response Assessments:

Central review performed by a panel of independent, expert hematologists according to an approved charter under change control (primary response assessment); local review of bone marrow and blood test results by the investigator at the study center where patient is enrolled (secondary assessment); “sponsor review” by employees and/or agents of Tolero Pharmaceuticals

Missing Values:

Missing values will not be replaced by imputed values. Patients who do not supply bone marrow for response assessment are counted among those not achieving remission. Missing time-to-event endpoints will be censored using standard rules detailed below.

Interim Analysis:

Single interim analysis after cycle-1 responder status is established for 28 patients per treatment arm; O’Brien-Fleming boundaries to control type-1 error rate

Significance Level:

One-sided, 2.5% including control for interim analysis and multiple comparisons

Procedure for Multiple Comparisons:

Hierarchical closed test procedure

Efficacy Endpoints:

Cycle-1 CR rate (primary endpoint), cycle-1 combined CR rate, cycle-1 combined remission rate, stem cell transplant rate, overall survival, progression-free survival, relapse-free survival, CR duration

Safety Endpoints:

30- and 60-day survival, adverse events, time to neutrophil recovery, time to platelet recovery, prevalence of tumor lysis syndrome, clinical laboratory parameters, vital signs

Statistical Methods:

Treatment-arm comparisons of cycle-1 complete remission rates (primary efficacy endpoint) and stem cell transplant rates by stratified CMH general association test (primary analysis) and logistic regression (supportive analyses); comparisons of time-to-event endpoints (progression-free survival, overall survival, relapse-free survival, and complete remission duration) by stratified log rank test (primary analysis) and Cox PH models (supportive analyses)

2. Abbreviations

Standard abbreviations (e.g., AML = acute myeloid leukemia and mg = milligrams) are omitted intentionally.

ACM	Alvocidib, Ara-C (Cytarabine) and Mitoxantrone
AE	Adverse Event
ANC	Absolute Neutrophil Count
ATC	Anatomical and Therapeutic Class
AUC	Area Under Curve
BM	Bone Marrow
BSA	Body Surface Area
CI	Confidence Interval
CM	Ara-C (Cytarabine) and Mitoxantrone
CMH	Cochran-Mantel-Haenszel (test)
CR	Complete Remission
CRi	Complete Remission with Incomplete Blood Count Recovery
CRp*	Complete Remission with Incomplete Platelet Recovery
CTCAE	Common Terminology Criteria for Adverse Events
DDE	Drug Dictionary Enhanced
DSMB	(Independent) Data Safety Monitoring Board
ECOG	Eastern Cooperative Oncology Group (performance status)
HR	Hazards Ratio
ITT	Intent-to-Treat (patient population)
IWG	International Working Group
KM	Kaplan-Meier (survivor function estimate)
LR	Logistic Regression
MedDRA	Medical Dictionary for Regulatory Activities
OS	Overall Survival
PBSC	Peripheral Blood Stem Cells
PFS	Progression-Free Survival
PH	Proportional Hazards (model)

* The term CRp appears in this document to maintain compatibility with the study's case report forms, however CRp will not be used by the hematologists performing the central review of patients' responses and it is not expected to be used by the investigators.

PR	Partial Remission
PT	(MedDRA) Preferred Term
RFS	Relapse-Free Survival
ROC	Receiver Operating Characteristic (curve)
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCT	Stem Cell Transplant/Transplantation
SMQ	Standardized MedDRA Query
SOC	(MedDRA) System Organ Class
TEAE	Treatment Emergent Adverse Event
TLS	Tumor Lysis Syndrome
WHO	World Health Organization

3. Definitions

Standard statistical terminology (e.g., type-1 error rate and significance level) is omitted intentionally. Efficacy endpoints are defined in Section 4.1 below (see Table 1).

Baseline Value: last measured value prior to the first dose of study drug (or prior to randomization for patients not treated)

Combined CR: response assessment of CR, CRi or CRp

Combined Remission: response assessment of CR, CRi, CRp or PR

Crossover Patient: a patient randomized to CM who, following 1-2 cycles of CM, receives ACM as stipulated in the study protocol

Cycle 1: period of time from the first dose of study drug until the start of study drug in cycle 2 or the start of alternative AML therapy other than supportive measures such as blood product transfusions (hence a patient whose first response assessment in cycle 1 is a CRi or CRp is considered to have achieved CR if neutrophils and platelets recover prior to initiation of cycle 2 or alternative AML therapy)

Cycle-1 Combined CR Rate: percentage of patients who achieve a combined CR (CR, CRi or CRp) after one cycle of treatment

Cycle-1 Combined Remission Rate: percentage of patients who achieve a combined remission (CR, CRi, CRp or PR) after one cycle of treatment

Cycle-1 CR Rate: percentage of patients who achieve a CR after one cycle of treatment (see section 7.9.1)

Intent-to-Treat Population: all patients randomized to receive study drug regardless of whether they actually received study drug and regardless of whether evidence is found indicating they failed to meet study inclusion/exclusion criteria or had other protocol violations (see section 7.1)

Neutrophil Recovery: Absolute neutrophil count ≥1x10^9/L following the administration of study drug

Per-Protocol Population: The subset of safety patients who either (a) have at least one response assessment on study, or (b) die prior to their first scheduled response assessment; however, patients are excluded from the per-protocol population if they have an important protocol deviation, such as when evidence is found indicating they failed to meet any study inclusion/exclusion criterion (see section 7.1)

Platelet Recovery: Platelet count ≥100x10^9/L following the administration of study drug

Pointwise: without controlling for multiple comparisons

Prestudy: Occurrence prior to randomization

Response-Available Population: The subset of safety patients who have at least one response assessment on study (see section 7.1)

Safety Population: The subset of intent-to-treat patients who received at least one dose of study drug (see section 7.1)

Sensitivity Analysis: study of how statistical methodology and assumptions affect results of the primary analysis

Shift Table: two-way frequency table pairing baseline value with the most extreme post-baseline result

Study Drug: alvocidib, Ara-C (cytarabine) and mitoxantrone

Supportive Analysis: synonymous with Sensitivity Analysis

Treatment Failure: failing to achieve a CR, CRi, CRp or PR following induction therapy (see section 7.9.2)

4. Introduction

This Statistical Analysis Plan (SAP) details the planned methodology for summarizing data collected in the clinical study conducted under Tolero, Inc. protocol TPI-ALV-201 and for comparing outcomes between the two treatment arms, ACM (alvocidib, Ara-C [cytarabine] and mitoxantrone) and CM (Ara-C [cytarabine] and mitoxantrone). The current version of the study protocol is Amendment 7 dated 30-MAY-2018. Future protocol amendments (if any) will be reviewed to assess whether the changes necessitate modification of this SAP.

Since the study protocol is a companion document to this SAP, aspects in the protocol unrelated to statistical issues (e.g., patient eligibility criteria, descriptions of clinical materials, and criteria for defining treatment failure, response to treatment and disease relapse) are not repeated here.

4.1. Study Endpoints

Statistical analysis and/or summarization is planned for the following study endpoints. Complete details on how endpoints are calculated and compared across treatment arms are presented in the subsections below. A key to abbreviated terms appears in Section 2.

Table 1. Study Endpoints

Primary Efficacy Endpoint • Cycle-1 CR Rate: percentage of patients who achieve a CR after one cycle of treatment
Secondary Efficacy Endpoints (not listed in order of importance) • Cycle-1 Combined CR Rate: percentage of patients achieving a combined CR (CR, CRi or CRp) after one cycle of treatment • Cycle-1 Combined Remission Rate: percentage of patients achieving a combined remission (CR, CRi, CRp or PR) after one cycle of treatment • Overall Survival: time from the date of randomization (Day 1) until death from any cause • Progression-Free Survival: time from the date of randomization (Day 1) until (a) treatment failure, (b) relapse after combined remission (CR, CRi, CRp or PR), or (c) death from any cause, whichever occurs first • Relapse-Free Survival Following CR (defined only for patients who achieve a CR during the study): time from the date of CR (Day 1) until (a) relapse after CR, or (b) death from any cause, whichever occurs first • Relapse-Free Survival Following Combined CR (defined only for patients who achieve a combined CR during the study): time from the earliest date of combined CR (Day 1) until (a) relapse after combined CR, or (b) death from any cause, whichever occurs first • Relapse-Free Survival Following Combined Remission (defined only for patients who achieve a combined remission during the study): time from the earliest date of combined remission (Day 1) until (a) relapse after combined remission, or (b) death from any cause, whichever occurs first • CR Duration (defined only for patients who achieve a CR during the study): time from the date of CR (Day 1) until relapse after CR • Combined CR Duration (defined only for patients who achieve a combined CR during the study): time from the earliest date of combined CR (Day 1) until relapse after combined CR • Combined Remission Duration (defined only for patients who achieve a combined remission during the study): time from the earliest date of combined remission (Day 1) until relapse after combined remission • Stem Cell Transplant Rate: percentage of patients proceeding directly to stem cell transplantation within 90 days after their last dose of study drug

Table 1. Study Endpoints

Safety Endpoints (not listed in order of importance) • 30- and 60-Day Mortality • Adverse Event Incidence Rates • Time to neutrophil recovery ≥1x10^9/L • Time to platelet recovery ≥100x10^9/L • Presence and Severity of Tumor Lysis Syndrome • Clinical Laboratory Parameters • Vital Signs
Other Endpoints • Concomitant Medication Usage Rates

4.2. Statistical Objectives

The statistical objectives are fourfold:

1. To establish the statistical significance at a controlled, one-sided 2.5% significance level, the superiority of the ACM cycle-1 CR rate in relapsed/refractory AML patients with baseline MCL-1 dependence ≥40% compared to that for CM treatment
2. To demonstrate the statistical significance of the superiority of ACM treatment compared to CM with respect to secondary measures of efficacy while controlling for the inflationary effects of multiple comparisons on the overall significance level
3. To model the relationship between baseline MCL-1 dependence and response to treatment with ACM
4. To describe the safety of ACM treatment in relapsed/refractory and newly diagnosed, high-risk AML patients (safety in relapsed/refractory AML patients will be judged relative treatment with CM)

5. Sample Size

5.1. Primary Study

5.1.1. Stage 1

The sample size of 23 eligible patients for stage 1 was selected because it allows estimation of the CR rate by means of a 90% confidence interval with maximum width of ±17%. Although this study does not follow a Simon 2-stage design, 23 patients is consistent with the sample size for stage 1 of the Simon 2-stage minimax design with 80% power for testing the null hypothesis that the CR rate is 50% against a one-sided alternative at the 5% level of significance when the actual CR rate is 70%. In the Simon

2-stage design described, the study would continue to stage 2 only if 13 patients achieve remission out of a maximum of 23 eligible patients. Therefore, this was selected as the requirement to proceed to stage 2 of this study.

5.1.2. Stage 2

The total planned sample size of 106 patients (53 per treatment arm) in stage 2 of the primary study provides 90% power to reject the null hypothesis of equal cycle-1 CR rates across treatment arms under the following conditions:

- Cycle-1 CR rates of 70% and 40% for patients receiving ACM and CM, respectively
- One-sided 2.5% significance level
- A single interim analysis of cycle-1 CR rates using O'Brien-Fleming boundaries[1] (see section 7.5.1) after responder status has been defined for 56 patients (28 per treatment arm)
- Randomization based on a 1:1 patient allocation ratio (ACM to CM)

Table 2 lists the programming statements used in SAS (version 9.4) to compute the sample size. Relevant SAS output is shown in Table 3 and Table 4.

Table 2. SAS Programming Statements for Sample Size Calculation

```
proc seqdesign boundaryscale=pvalue;
   design alpha=0.025 alt=upper beta=0.1 info=cum(56 106) method=obf nstages=2;
   samplesize model=twosamplefreq(nullprop=0.40 prop=0.70 test=prop weight=1);
run;
```

Table 3. SAS PROC SEQDESIGN Output: Boundary Information

				Alternative	Boundary Values
	Information Level			Reference	Upper
Stage	Proportion	Actual	N	Upper	Alpha
1	0.5283	62.2073	55.98657	2.36615	0.00321
2	1.0000	117.7495	105.9746	3.25537	0.02378

Table 4. SAS PROC SEQDESIGN Output: Sample Sizes

	Fractional N				Ceiling N			
Stage	N	N(Grp 1)	N(Grp 2)	Information	N	N(Grp 1)	N(Grp 2)	Information
1	55.99	27.99	27.99	62.2073	56	28	28	62.2222

Table 4. SAS PROC SEQDESIGN Output: Sample Sizes

	Fractional N				Ceiling N			
Stage	N	N(Grp 1)	N(Grp 2)	Information	N	N(Grp 1)	N(Grp 2)	Information
2	105.97	52.99	52.99	117.7	106	53	53	117.8

5.2. Exploratory Study Arms

The sample size of 20 eligible patients in each of the four exploratory arms was selected because it allows estimation of the CR rate by means of a 90% confidence interval with maximum width of ±18%.

6. Treatment Randomization

All patients participating in stage 1 of the primary study or an exploratory study arm will be assigned treatment with ACM. Randomization is not used in these parts of the study.

The remainder of this section refers only to patient participating in stage 2 of the primary study.

Stage 2 of the primary study follows a randomized, two-arm parallel design. The two treatment arms are ACM and CM. Dosages and infusion times are detailed in the study protocol.

Enrolled patients are randomized to receive ACM or CM in a 1:1 ratio and stratified by a three-category classification of response to first-line therapy:

- Refractory – defined as persistent disease or a CR duration less than 90 days
- Early Relapse – defined as a CR duration between 90 days and 1 year
- Late Relapse – defined as a CR duration greater than 1 year but less than 24 months

7. Statistical Methods

Statistical inference (hypothesis testing) to address the first and second objectives shown in section 4.2 will be addressed using data from stage 2 of the primary study only. Data from stage 1 of the primary study and the exploratory study arms will not be used for statistical inference.

Analyses will be performed using SAS statistical software. Default estimation methods in version 9.4 of SAS are always used unless an alternative is specified below. Data summaries will include the mean, standard deviation, median, minimum and maximum values for continuous data; the median, 25^{th} and 75^{th} percentiles, minimum and maximum values for time-to-event endpoints; and the number and percentage of patients in each category for categorical data. Pointwise 95% confidence intervals (CI) will also be estimated for the mean (continuous data), median (time-to-event endpoints) or percentage of patients (categorical data).

In general, data values will be obtained directly from the clinical database (e.g., birth date, treatment dates, AML relapse date and death date). Derivations for certain values not collected explicitly in the database are explained in relevant sections below (e.g., age and treatment failure date, sections 7.7 and 7.9.2, respectively). In some instances, it is beyond the scope of this SAP to explain schemes for patient classifications and data derivations (e.g., classifying patients in cytogenetic risk group according to published criteria). These details will be expounded in the clinical study report.

Durations of time expressed in months (e.g., times-to-event endpoints) will be reported in proportional "months" according to the formula

$$\text{months} = \frac{12}{365.25} \times \text{days}$$

Baseline value of a characteristic is defined as the last measured value prior to the first dose of study drug (or prior to randomization for patients not treated).

Statistical p-values will be truncated for display at the third digit following the decimal point. P-values less than 0.001 will be displayed as "<0.001."

7.1. Analysis Populations

The intent-to-treat (ITT) patient population includes all patients randomized to receive study drug regardless of whether they actually received study drug and regardless of whether evidence is found indicating they failed to meet study inclusion/exclusion criteria or had other protocol violations. When the ITT patient population is analyzed, patients are grouped according to their randomized treatment regardless of actual treatment received. The ITT patient population is the analysis population for the primary analyses of efficacy endpoints (i.e., all statistical inference and tests of hypotheses related to efficacy endpoints in stage 2 and univariate summaries of efficacy endpoints in stage 1 and the exploratory study arms).

The safety patient population is the subset of ITT patients who received at least one dose of study drug. When the safety patient population is analyzed, patients are grouped according to actual treatment received. ACM is considered a patient's "treatment received" in the safety patient population if alvocidib is administered at any time during the study (except for ACM administered to a "crossover" patient). The safety patient population is the analysis population for all analyses of safety data.

The response-available patient population is the subset of safety patients who have at least one response assessment on study. However, patients are excluded from the response-available patient population if they have an important protocol deviation, such as when evidence is found indicating they failed to meet any study inclusion/exclusion criterion. When the response-available patient population is analyzed, patients are grouped according to actual treatment received. ACM is considered a patient's "treatment received" in the response-available patient population if alvocidib is administered at any time prior to a combined remission or at any time during the study for patients never achieving combined remission (except for ACM administered to a "crossover" patient). The response-available population is used for biomarker analyses and sensitivity (secondary, supportive) analyses of efficacy endpoints.

The per-protocol patient population is the subset of safety patients who either (a) have at least one response assessment on study, or (b) die prior to their first scheduled response assessment. However, patients are excluded from the per-protocol patient population if they have an important protocol deviation, such as when evidence is found indicating they failed to meet any study inclusion/exclusion criterion. When the per-protocol patient population is analyzed, patients are grouped according to actual treatment received. ACM is considered a patient's "treatment received" in the per-protocol patient population if alvocidib is administered at any time prior to a combined remission or at any time during the study for patients never achieving combined remission (except for ACM administered to a "crossover" patient). The per-protocol population is used for sensitivity (secondary, supportive) analyses of efficacy endpoints.

The crossover patient population includes all patients randomized to CM who, following 1-2 cycles of CM, receive ACM as stipulated in the study protocol.

Each of the patient populations listed above may be further subdivided for certain analyses. For example, data from ITT patients with relapsed/refractory AML who received ACM (i.e., the ITT population excluding relapsed/refractory patients receiving CM in cycle 1 and patients with newly diagnosed, high-risk AML) will be analyzed to quantify the relationship between baseline MCL-1 dependence and the independent binary variable cycle-1 CR (see section **Error! Reference source not found.**).

7.2. Missing Values

In general, missing values will not be replaced by imputed values.

- Although the central hematologists' assessment of response is the primary evaluation, the local assessment (i.e., the one recorded at the study center where the patient is enrolled) will be used for patients who are missing an evaluation from the central hematologists (see section 7.9.1)
- Patients who, for any reason, do not supply bone marrow for response assessment are counted among those not achieving remission
- Patients missing a time-to-event endpoint will have their value censored according to the rules set forth below (see sections 7.9.2 to 7.9.6)

Exceptions to the general rule of not replacing missing values are planned for the following items.

- The first day of the month will be used for a missing day component of a partial birthdate, and January will be substituted for a missing birth month (these substitutions are relevant only for calculating a patient's age)
- Partial start and end dates for adverse events will be replaced by calendar dates that maximize the duration of the adverse event (see section 7.11.2)
- Missing clinical laboratory parameters and vital signs will be replaced using the "last value carried forward" method (see sections 7.11.4 and 7.11.5, respectively)

7.3. Visit Windows

Visit windows are not relevant for analyzing efficacy endpoints since these endpoints are either binary outcomes over the entire study period (e.g., CR rate) or measured on a continuous time scale (e.g., OS and PFS). Some safety assessments (clinical laboratory parameters and vital signs) are summarized over time and visit windows will be used (see sections 7.11.4 and 7.11.5, respectively).

7.4. Interim Analysis

A single interim analysis is planned when cycle-1 response has been defined for 28 patients per treatment arm. Cycle-1 response may be available for more than 28 patients in one treatment arm when the alternate treatment arm reaches 28 because the enrollment ratio may not be precisely 1:1 over time. The additional patients in the faster-enrolling treatment arm will be included in the interim analysis. Therefore, the interim analysis sample size may be slightly more than 56 patients.

Cycle-1 CR rate (the primary efficacy endpoint) will be tested using the statistical procedures described below. As shown in Table 3, a p-value ≤0.00321 is required to reject the null hypothesis of equal cycle-1 CR rates across treatment arms.

Measures taken to counter the effect of the interim analysis and maintain an overall 2.5% significance level are presented in section 7.5.1.

7.4.1. Outcomes and Study Continuation

Two outcomes are possible from the interim analysis: statistical significance or nonsignificance. In the event of a nonsignificant difference between ACM and CM Cycle-1 CR rates (p-value >0.00321), the study will continue to accrue and randomize patients.

If the interim analysis result for the primary analysis of Cycle-1 CR rate (see section 7.9.1) is statistically significant (p-value ≤0.00321), then randomization will be stopped. Additional eligible stage-2 patients will be enrolled into the study to receive open-label ACM until the planned enrollment of 53 ACM patients is complete. No additional patients will be enrolled into the CM treatment arm. Terminating randomization after the interim analysis will have a negative effect on the power to declare treatment differences in secondary endpoints as statistically significant, however it is understood from polling investigators that they would be unwilling to enroll patients into the study once ACM has been established as superior to CM with respect to Cycle-1 CR rate.

7.5. Overall Type-1 Error Rate

A type-1 error in statistics is rejecting the null hypothesis as false when it is actually true. In this study, this corresponds to concluding the population's efficacy response following ACM is superior to that from CM when in fact it is not. The probability of committing a type-1 error is commonly referred to as the significance level or α (alpha) level. The overall significance level for this study will be 2.5% using one-sided hypothesis tests.

7.5.1. Type-1 Error Rate Control for Interim Analysis

An O'Brien-Fleming sequential design[1] is employed to maintain the overall 2.5% significance level of the study following the single interim analysis described in section 7.4. Although the interim analysis will include several safety endpoints, cycle-1 CR rate will be the only efficacy endpoint considered when determining whether O'Brien-Fleming boundaries have been crossed. Programming statements in Table 2 were run in SAS software (version 9.4) giving output in Table 3 and Table 4. Thus, according to the O'Brien-Fleming boundaries, the final analysis will use a one-sided significance level of 2.378% to determine statistical significance (see Table 4).

7.5.2. Type-1 Error Rate Control for Multiple Comparisons

Without taking appropriate measures, testing multiple efficacy parameters inflates the overall significance level since the probability of making a type-1 error in at least one from a set of hypothesis tests is greater than the probability of making a type-1 error in any single test. The hierarchical closed test procedure[3,4] will be used to maintain the overall 2.5% significance level.

The order of efficacy endpoints for the hierarchical closed test procedure is:

1. Cycle-1 CR rate
2. Progression-Free Survival
3. Stem Cell Transplant Rate
4. Cycle-1 Combined CR Rate
5. Cycle-1 Combined Remission Rate
6. Overall Survival
7. Combined Remission Duration
8. Combined CR Duration
9. CR Duration

The final analysis will be conducted at a one-sided 2.378% significance level using the statistical procedures described below. Cycle-1 CR rate, the primary efficacy endpoint, will be declared statistically significant if its test p-value ≤0.02378. A secondary efficacy endpoint will be declared statistically significant only if its test p-value ≤0.02378 and all efficacy endpoints preceding it in the hierarchy are statistically significant. In other words, declaration of statistical significance for efficacy endpoints will advance in order down the hierarchy only until a p-value >0.02378 is found.

In addition, although primary and supportive analytical methods are specified for each endpoint in the hierarchy, only the p-value from the primary analysis will be used for deciding statistical significance. Supportive (sensitivity) analyses are to demonstrate the robustness of primary results to slight modifications to how endpoints are defined and/or analyzed.

Declarations of statistical significance will not be advanced for endpoints not included in the hierarchy of the closed test procedure (e.g., ECOG performance status).

It is worthwhile noting…

The testing hierarchy is not intended to represent a ranking of efficacy endpoints based solely on clinical importance. It resulted from a careful consideration balancing clinical importance, perceived regulatory acceptability, statistical power (presumed probability of treatment-arm differences achieving statistical significance), and interrelatedness of the efficacy endpoints. This study is not powered to demonstrate statistical significance of OS differences typically observed in clinical studies involving relapsed/refractory AML patients. None of this however would invalidate or undermine the extreme clinical importance of a statistically significant OS advantage for ACM over CM observed in this study.

7.6. Patient Disposition

The numbers of patients randomized (ITT population), dosed (safety population) and included in the response-available, per-protocol and crossover populations will be summarized by frequency counts within treatment arm. Time on study will be summarized as a continuous variable within treatment arm (mean, standard deviation, median, minimum and maximum values). Reasons for discontinuation will be summarized within treatment arm as a categorical variable (number and percentage of patients in each category). Relevant data supporting these summaries will be listed by patient within treatment arm.

7.7. Baseline Characteristics

Patient age in years on the date of randomization will be calculated by dividing the number of days since birth (Day 1) by 365.25. If body surface area (BSA) expressed in m^2 is missing in the study database, then it will be calculated using the formula[5]

$$\text{BSA} = \sqrt{\frac{W \times H}{3600}}$$

where W is the patient's weight in kg and H is the patient's height in cm.

Summaries of baseline characteristics will include within– and between–treatment-arm descriptive statistics appropriate for continuous, time-to-event and categorical variables (see section 7). Separate summaries will be produced for the ITT, safety, response-available, per-protocol and crossover patient populations. The Student *t* test (continuous variables), the log rank test (time-to-event variables), or the Fisher exact test (categorical variables) will be used to quantify the effectiveness of randomization to create similar distributions of these variables between treatment arms. The effect any baseline characteristic with two-sided p-value <0.10 has on efficacy outcomes will be investigated (see sections 7.9.1 through 7.9.6).

Baseline characteristics include demographic data (e.g., age, sex, race, ethnicity, height, weight and BSA); AML history (e.g., clinical onset of AML [de novo AML, prior MDS or prior leukemogenic therapy], time since AML diagnosis, peripheral blood and bone marrow blast percentages, cytogenetic risk group, frontline AML therapy and response to frontline AML therapy [see section 6]); MCL-1 dependence; and other

medical history (e.g., baseline ECOG performance status, hemoglobin, leukocyte and platelet counts, and blood chemistry parameters assessing liver and kidney function). Reported medical history terms will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) preferred terminology and then summarized within and between treatment-arms by incidence rates for each MedDRA primary system organ class (SOC) and preferred term (PT). Optionally, depending on the perceived relevance, prestudy medications will mapped to terminology in the World Health Organization (WHO) Drug Dictionary Enhanced (DDE) and then summarized within treatment arm by usage rates for each WHO DDE anatomical and therapeutic class (ATC) and preferred (i.e., standardized) drug name. Relevant data supporting these summaries will be listed by patient within treatment arm.

7.8. Study Drug Exposure

Patients' total exposures to study drugs and average exposures per treatment cycle will be calculated and summarized within treatment arm by the mean, standard deviation, median, minimum and maximum values. Numbers of treatment cycles received will be summarized within treatment arm as a continuous variable (by the mean, standard deviation, median, minimum and maximum values) and as a categorical variable (by the number and percentage of patients in each category). Reasons for noncompliance with protocol-specified dosing schedules will be summarized within treatment arm as a categorical variable. Relevant data supporting these summaries (including actual infusion durations and expected doses based on BSA calculated from a patient's latest reported height and weight) will be listed by patient within treatment arm.

7.9. Efficacy Endpoints

Relevant data supporting efficacy analyses will be listed by patient within treatment arm.

7.9.1. Remission Rates

The primary efficacy endpoint is cycle-1 CR rate (i.e., the percentage of patients who achieve a CR after one cycle of treatment). Cycle 1 does not end until the patient starts cycle 2 or alternative AML therapy other than supportive measures (e.g., blood product transfusions). Therefore, a patient whose first response assessment in cycle 1 is a CRi or CRp[*] is considered to have achieved CR if neutrophils and platelets recover prior to initiation of cycle 2 or alternative AML therapy including bone marrow or stem cell transplant. Patients who, for any reason, do not supply bone marrow for response assessment are counted among those not achieving remission. Best response during the study (i.e., best response following one or more cycles of treatment) was not selected as the primary efficacy endpoint due to the anticipated confounding effects of patients who are randomized to CM but crossover after one or two cycles to receive treatment with ACM.

Three response assessments are planned for each postbaseline bone marrow sample:

1. The "central review" performed by a panel of independent, expert hematologists according to an approved charter under change control

[*] See footnote on page 6.

2. The “local review” of bone marrow and blood test results by the investigator at the study center where patient is enrolled
3. The “sponsor review” by employees and/or agents of Tolero Pharmaceuticals

It is beyond the scope of this SAP to restate or summarize the written charter detailing the methods and criteria the panel of hematologists use to grade response, but both central and local assessments are made according to criteria specified in the study protocol. The primary assessments of remission for comparing efficacy across treatment arms is the central review of bone marrow; however, in cases where the central hematologists are unable to rate response, the investigator’s response assessment will be used in place of the missing central assessment. Supportive efficacy analyses will be conducted using assessments from the local and sponsor reviews.

Hypothesis testing will be based on the stratified Cochran-Mantel-Haenszel (CMH) general association test statistic[6,7] (see Table 5). Stratification levels will be the same as those used for the randomization scheme (see section 6). Although the alternative hypothesis for the CMH general association test statistic is that cycle-1 CR is correlated with treatment arm in at least one stratum, the statistic has low power for detecting an association in which the patterns of association for some of the strata are in the opposite direction of the patterns displayed by other strata.[8] Therefore, a statistically significant CMH general association statistic will validate the positive contribution of alvocidib (added to CM) to induce CR (over CM alone).

Within–treatment-arm statistics will include the number and percentage of patients rated as not achieving CR due to not providing a bone marrow biopsy/aspirate, proportion of patients who achieve a cycle-1 CR and exact 95% CI estimated without stratification. Between–treatment-arm statistics will include the Mantel-Haenszel odds ratio and 95% CI estimated with stratification.

A standard logistic regression (LR) model will be estimated relating the probability of a cycle-1 CR as a function of baseline MCL-1 dependence, treatment arm, and the two-way interaction between baseline MCL-1 dependence and treatment arm. If the p-value for the Wald chi-square statistic for two-way interaction is >0.10, then the Wald chi-square statistic for treatment arm will be interpreted as testing the equality of cycle-1 CR rates controlling for the effect of baseline MCL-1 dependence. Results will be presented graphically as follows. Patients will be divided into groups according to which range of values their baseline MCL-1 dependence falls:

$$(0\%\text{-}\leq d_1),\ (d_1\text{-}\leq d_2),\ (d_2\text{-}\leq d_3),\ \ldots,\ (d_8\text{-}\leq d_9),\ (d_9\text{-}100\%)$$

where $d_1,\ldots,d_9$ are deciles for baseline MCL-1 dependence regardless of treatment arm. Empirical within–treatment-arm cycle-1 CR rates will be calculated for patients within each range and plotted alongside cycle-1 CR probabilities estimated from the LR model.

Table 5. Statistical Analyses of Cycle-1 CR Rates

Analysis Population	Analytical Method	Response Assessment	Role of Analysis
Intent-to-Treat	Stratified CMH	Central	Primary
Intent-to-Treat	Stratified CMH	Local and Sponsor	Supportive

Table 5. Statistical Analyses of Cycle-1 CR Rates

Analysis Population	Analytical Method	Response Assessment	Role of Analysis
Intent-to-Treat	Logistic regression	Central	Supportive
Intent-to-Treat	Logistic regression	Local and Sponsor	Supportive
Per-Protocol	Stratified CMH	Central	Supportive
Per-Protocol	Stratified CMH	Local and Sponsor	Supportive
Per-Protocol	Logistic regression	Central	Supportive
Per-Protocol	Logistic regression	Local and Sponsor	Supportive
Response-Available	Logistic regression	Central	Supportive
Response-Available	Logistic regression	Local and Sponsor	Supportive
Subgroups based on patient characteristics	Logistic regression	Central, Local and Sponsor	Exploratory

If the test for between–treatment-arm homogeneity of any baseline characteristic gives two-sided p-value <0.10, then the effect that baseline characteristic has on cycle-1 CR rate will be quantified using LR. Two LR models will be fit. One model will include the baseline characteristic in question as the only independent variable, and the second model will include the baseline characteristic and the stratification factors used for randomization.

Other remission rates (e.g., cycle-1 combined CR rate and cycle-1 combined remission rate) will be summarized similar to cycle-1 CR rate.

Correlation between central and local review will be assessed descriptively in 2-way frequency tables, both as a bivariate response (CR yes or no) and on the full scale (CR, CRi, CRp, PR, resistance/relapse). The simple kappa coefficient[8,9] will be calculated as a measure of agreement between the central and local reviews. Since the two response variables are two independent ratings of the same tissue sample, the kappa coefficient equals 1 when there is complete agreement of the raters. When the observed agreement exceeds chance agreement, kappa is positive, with its magnitude reflecting the strength of agreement. Although this is unusual in practice, kappa is negative when the observed agreement is less than chance agreement.[8]

7.9.2. Progression-Free Survival

PFS time is defined for all patients and is measured from the date of randomization (Day 1) until (a) treatment failure, (b) relapse after combined remission, or (c) death from any cause, whichever occurs first.

Treatment failure is defined as failing to achieve a combined remission following induction therapy. The date of treatment failure is the protocol-specified date for bone marrow collection following the first induction attempt (i.e., at neutrophil and platelet recovery or Day 45 of the treatment cycle, whichever occurs first), regardless of whether the patient received a second induction attempt (unless the second induction resulted in a combined remission in which case “treatment failure” does not apply).

Additional rules for defining PFS time are:

- A patient found to have relapsed after missing one or more scheduled response assessments will have his PFS time end on the scheduled date for the first missed response assessment following the last confirmation of combined remission
- A patient who misses the first scheduled response assessment and who never achieves combined remission will have his PFS time end on the scheduled date for the first response assessment (unless the patient died before this date)

Rules for censoring PFS time are:

- A patient with PFS continuing as of the date he is lost to follow-up will have his PFS time censored on the date of his last response assessment
- A patient with PFS continuing as of the database lock will have his PFS time censored on the date of his last response assessment

Hypothesis testing for PFS will be based on the stratified log rank test statistic (see Table 6). Stratification levels will be the same as those used for the randomization scheme (see section 6). Two Cox proportional hazards (PH) models will be fit to provide estimates of the hazards ratio (HR). One model will include treatment arm as the only independent variable, and the second model will include treatment arm and the stratification factors used for randomization. Kaplan-Meier (KM) "survival" curves[10] will be graphed for visual inspection of differences in PFS between the treatment arms. Within–treatment-arm statistics will include the number and percentage of censored PFS times, quartile PFS times, 95% CIs for the quartile PFS times, and KM "survival" percentages with 95% CIs at select milestone times (e.g., 6, 12 and 18 months). Between–treatment-arm statistics will include HRs from the Cox PH models and their 95% CIs.

Sensitivity analyses will be run with PFS time censored on the date of stem cell transplantation (bone marrow [BM] or peripheral blood stem cells [PBSC]).

Table 6. Statistical Analyses of PFS, OS, RFS and CR Duration

Analysis Population	Analytical Method	Censor for Transplant	Role of Analysis
Intent-to-Treat	Stratified log rank	No	Primary
Intent-to-Treat	Stratified log rank	Yes	Supportive
Intent-to-Treat	Cox PH Model	No	Supportive
Intent-to-Treat	Cox PH Model	Yes	Supportive
Per-Protocol	Stratified log rank	No	Supportive
Per-Protocol	Stratified log rank	Yes	Supportive
Per-Protocol	Cox PH Model	No	Supportive
Per-Protocol	Cox PH Model	Yes	Supportive
Subgroups based on patient characteristics	Cox PH model	No	Exploratory

If the test for between–treatment-arm homogeneity of any baseline characteristic gives two-sided p-value <0.10, then the effect that baseline characteristic has on PFS time will be quantified using Cox PH modeling. Two PH models will be fit. One model will include the baseline characteristic in question as the only independent variable, and the second model will include the baseline characteristic and the stratification factors used for randomization.

7.9.3. Stem Cell Transplant Rate

Stem cell transplant (SCT) rate is the percentage of patients proceeding to SCT within 90 days after their last dose of study drug regardless of whether the patient remains enrolled in this study. Unless documentation of SCT is provided, patients who are lost to follow-up within the 90 days following their last dose of study drug are counted among those not receiving SCT.

Hypothesis testing will be based on the stratified CMH general association test statistic (see Table 7). Stratification levels will be the same as those used for the randomization scheme (see section 6).

Within–treatment-arm statistics will include the number and percentage of patients counted as not receiving SCT due to being lost to follow-up; proportion of patients receiving SCT within 90 days after their last dose of study drug; and exact 95% CI estimated without stratification. Between–treatment-arm statistics will include the Mantel-Haenszel odds ratio and 95% CI estimated with stratification.

Table 7. Statistical Analyses of Stem Cell Transplant Rate

Analysis Population	Analytical Method	Role of Analysis
Intent-to-Treat	Stratified CMH	Primary
Per-Protocol	Stratified CMH	Supportive
Subgroups based on patient characteristics	Logistic regression	Exploratory

If the test for between–treatment-arm homogeneity of any baseline characteristic gives two-sided p-value <0.10, then the effect that baseline characteristic has on SCT rate will be quantified using LR. Two LR models will be fit. One model will include the baseline characteristic in question as the only independent variable, and the second model will include the baseline characteristic and the stratification factors used for randomization.

7.9.4. Overall Survival

OS time is defined for all patients and is measured from the date of randomization (Day 1) until death from any cause.

Rules for censoring OS time are:

- A patient with OS continuing as of the date he is lost to follow-up will have his OS time censored on the date of his last assessment of any type
- A patient with OS continuing as of the database lock will have his OS time censored on the date of his last assessment of any type

Hypothesis testing for OS will be based on the stratified log rank test statistic (see Table 6). Stratification levels will be the same as those used for the randomization scheme (see section 6). Two Cox PH models will be fit to provide estimates of the HR. One model will include treatment arm as the only independent variable, and the second model will include treatment arm and the stratification factors used for randomization. KM survival curves will be graphed for visual inspection of differences in OS between the treatment arms. Within–treatment-arm statistics will include the number and percentage of censored OS times, quartile OS times, 95% CIs for the quartile OS times, and KM survival percentages with 95% CIs at select milestone times (e.g., 6, 12 and 18 months). Between–treatment-arm statistics will include HRs from the Cox PH models and their 95% CIs.

Sensitivity analyses will be run with OS time censored on the date of BM or PBSC transplant.

If the test for between–treatment-arm homogeneity of any baseline characteristic gives two-sided p-value <0.10, then the effect that baseline characteristic has on OS time will be quantified using Cox PH modeling. Two PH models will be fit. One model will include the baseline characteristic in question as the only independent variable, and the second model will include the baseline characteristic and the stratification factors used for randomization.

7.9.5. Relapse-Free Survival Following CR

RFS time is defined only for patients who achieve a CR during the study and is measured from the date of CR (Day 1) until (a) relapse after CR, or (b) death from any cause, whichever occurs first.

An additional rule for defining RFS time is:

- A patient found to have relapsed after missing one or more scheduled response assessments will have his RFS time end on the scheduled date for the first missed response assessment following the last confirmation of CR

Rules for censoring RFS time are:

- A patient with RFS continuing as of the date he is lost to follow-up will have his RFS time censored on the date of his last response assessment
- A patient with RFS continuing as of the database lock will have his RFS time censored on the date of his last response assessment

Hypothesis testing for RFS will be based on the stratified log rank test statistic (see Table 6). Stratification levels will be the same as those used for the randomization scheme (see section 6). Two Cox PH models will be fit to provide estimates of the HR. One model will include treatment arm as the only independent variable, and the second model will include treatment arm and the stratification factors used for randomization. KM “survival” curves will be graphed for visual inspection of differences in RFS between the treatment arms. Within–treatment-arm statistics will include the number and percentage of patients who achieve a CR, the number and percentage of censored RFS times, quartile RFS times, 95% CIs for the quartile RFS times, and KM “survival” percentages with 95% CIs at select milestone times (e.g., 6, 12 and 18 months).

Between–treatment-arm statistics will include HRs from the Cox PH models and their 95% CIs.

Sensitivity analyses will be run with RFS time censored on the date of BM or PBSC transplant.

If the test for between–treatment-arm homogeneity of any baseline characteristic gives two-sided p-value <0.10, then the effect that baseline characteristic has on RFS time will be quantified using Cox PH modeling. Two PH models will be fit. One model will include the baseline characteristic in question as the only independent variable, and the second model will include the baseline characteristic and the stratification factors used for randomization.

Two endpoints analogous to relapse-free survival following CR will be analyzed using similar methods.

- Relapse-Free Survival Following Combined CR (defined only for patients who achieve a combined CR during the study): time from the earliest date of combined CR (Day 1) until (a) relapse after combined CR, or (b) death from any cause, whichever occurs first
- Relapse-Free Survival Following Combined Remission (defined only for patients who achieve combined remission during the study): time from the earliest date of combined remission (Day 1) until (a) relapse after combined remission, or (b) death from any cause, whichever occurs first

Response categories included in “combined CR” and “combined remission” are delineated in Section 3.

7.9.6. Complete Remission Duration

CR duration is defined only for patients who achieve a CR during the study and is measured from the date of CR (Day 1) until relapse.

An additional rule for defining CR duration is:

- A patient found to have relapsed after missing one or more scheduled response assessments will have his CR duration end on the scheduled date for the first missed response assessment following the last confirmation of CR

Rules for censoring CR duration are:

- A patient who dies while in remission will have his CR duration censored on the date of death
- A patient in remission on the date he is lost to follow-up will have his CR duration censored on the date of his last response assessment
- A patient in remission on the database lock will have his CR duration censored on the date of his last response assessment

Hypothesis testing for CR duration will be based on the stratified log rank test statistic (see Table 6). Stratification levels will be the same as those used for the randomization scheme (see section 6). Two Cox PH models will be fit to provide estimates of the HR. One model will include treatment arm as the only independent variable, and the second

model will include treatment arm and the stratification factors used for randomization. KM “survival” curves will be graphed for visual inspection of differences in CR duration between the treatment arms. Within–treatment-arm statistics will include the number and percentage of patients who achieve a CR, the number and percentage of censored CR durations, quartile CR durations, 95% CIs for the quartile CR durations, and KM “survival” percentages with 95% CIs at select milestone times (e.g., 6, 12 and 18 months). Between–treatment-arm statistics will include HRs from the Cox PH models and their 95% CIs.

Sensitivity analyses will be run with CR duration censored on the date of BM or PBSC transplant.

If the test for between–treatment-arm homogeneity of any baseline characteristic gives two-sided p-value <0.10, then the effect that baseline characteristic has on CR duration will be quantified using Cox PH modeling. Two PH models will be fit. One model will include the baseline characteristic in question as the only independent variable, and the second model will include the baseline characteristic and the stratification factors used for randomization.

Two endpoints analogous to CR duration will be analyzed using similar methods.

- Combined CR Duration (defined only for patients who achieve a combined CR during the study): time from the earliest date of combined CR (Day 1) until relapse after combined CR
- Combined Remission Duration (defined only for patients who achieve combined remission during the study): time from the earliest date of combined remission (Day 1) until relapse after combined remission

Response categories included in “combined CR” and “combined remission” are delineated in Section 3.

7.9.7. ECOG Performance Status

ECOG performance status will be summarized within treatment arm using shift tables.

7.10. Biomarker

The biomarker in this study is MCL-1 dependence measured by mitochondrial sensitivity to NOXA BH3 peptides. Relevant data supporting analyses of MCL-1 dependence will be listed by patient within treatment arm. Summary of baseline biomarker values will include within– and between–treatment-arm descriptive statistics appropriate for continuous variables (see section 7.7). An LR model will be fit to examine the relationship between baseline MCL-1 dependence and the independent binary variable cycle-1 CR for relapsed and refractory AML patients receiving ACM. In addition, area under the curve (AUC) will be calculated (by treatment arm) for the trapezoidal receiver operating characteristic (ROC) curve. This value will quantify the ability of MCL-1 dependence to predict cycle-1 CR.[11] Ninety-five percent CIs will be calculated by assuming estimated AUC follows a normal distribution.

$$\widehat{\text{AUC}} \pm z_{1-\alpha/2}\, SE_{\text{AUC}}$$

where the standard error SE_{AUC} is calculated by the jackknife method.[12] Data from relapsed/refractory patients receiving CM in cycle 1 and patients with newly diagnosed, high-risk AML will not be used in these analyses. Similar analyses will be conducted for (1) cycle-1 combined CR and (2) cycle-1 combined remission.

Quantifying the effect of MCL-1 dependence on response rates will not infer a safety and/or efficacy benefit (or detriment) of treatment with ACM over CM. Therefore, no statistical analyses of MCL-1 dependence across treatment arms are planned other than comparing baseline values to confirm randomization created homogeneous groups with respect to prestudy values.

7.11. Safety Endpoints

Summaries of safety endpoints will include data collected from the safety patient population. Relevant data supporting safety analyses will be listed by patient within treatment arm.

7.11.1. 30- and 60-Day Mortality

Estimates of 30- and 60-day mortality and their pointwise 95% CIs will be derived from KM curves for OS by treatment arm (see section 7.9.4).

7.11.2. Adverse Events

Reported adverse event (AE) terms will be mapped to MedDRA preferred terminology. AEs suggestive of tumor lysis syndrome (TLS) will be flagged based in the standardized MedDRA query (SMQ) for TLS. All reported events will appear in AE listings, however only treatment-emergent adverse events will be summarized. A treatment-emergent adverse event (TEAE) is an AE that starts or increases in severity any time after the first administration of any study drug up to 30 days following the last administration of any study drug. AE severity was rated by the investigator according to Common Terminology Criteria for Adverse Events (CTCAE) version 4.03 criteria.[13]

A high-level safety summary will display the numbers of patients within each treatment arm and overall who experience one or more AEs in each of the following categories:

- All TEAEs regardless of severity or presumed relationship to study drug
- TEAEs judged related to study drug
- All TLS SMQ TEAEs
- TLS SMQ TEAEs judged related to study drug
- Treatment-emergent serious adverse events (SAEs)
- TEAEs leading to a delay in the administration of study drug
- TEAEs leading to a reduction in the protocol-specified dose of study drug
- TEAEs leading to discontinuation of study drug
- TEAEs leading to withdrawal from the study
- TEAEs leading to death

The base summary of TEAEs will show within– and between–treatment-arm incidence rates for each MedDRA primary SOC and/or PT by highest reported CTCAE severity grade and overall. A separate summary will be produced each of the AE subsets listed above. Additional AE summaries may be produced using safety data from subsets of patients and/or characterized using additional SMQs.

Partial start and end dates for adverse events will be replaced by calendar dates that maximize the duration of the adverse event. The following steps will be followed.

For a partial start date:

1. Replace a missing month with January
2. Replace a missing calendar day with the first of the month
3. If the replacement date is prior to the first dose of study drug, then set the adverse event start date equal to the date of the first dose

For a partial end date:

1. Replace a missing month with December
2. Replace a missing calendar day with the last day of the month

The replacement date for a missing end date may exceed end of the adverse event reporting period specified in the study protocol.

An end date will not be estimated for adverse events marked continuing at the end of the study.

7.11.3. Classification of Tumor Lysis Syndrome

Cairo-Bishop definitions for laboratory TLS and clinical TLS will be used to diagnose and grade TLS.[14] Patients are classified as experiencing laboratory TLS if they have any two or more serum values of uric acid, potassium, phosphorus and/or calcium shown in Table 8 within 3 days before or 7 days after the initiation of study drug.

Table 8. Cairo-Bishop Definition of Laboratory Tumor Lysis Syndrome

Laboratory Parameter	Criteria
Uric Acid	≥476 µmol/L or 25% increase from baseline
Potassium	≥6.0 mmol/L or 25% increase from baseline
Phosphorus	≥1.45 mmol/L or 25% increase from baseline
Calcium	≤1.75 mmol/L or 25% decrease from baseline

Diagnosis of clinical TLS requires the presence of laboratory TLS and elevated serum creatinine levels, cardiac arrhythmia and/or seizures. Criteria for grading the severity of clinical TLS are presented in Appendix 1.

The numbers of patients experiencing laboratory TLS and the various severities of clinical TLS will be summarized within and between treatment arms.

7.11.4. Clinical Laboratory Parameters

Typical comparisons of results summarized by visit may have limited interpretability in this study because (1) study drugs from the two treatment arms are administered on

different cycle days (Days 1-9 for ACM versus Days 1-4 for CM), (2) the schedule for clinical laboratory evaluations (cycle day and frequency) is different between the treatment arms, and (3) additional cycles of treatment start according to patients' response to treatment rather than based on a predetermined schedule. Therefore, four groups of laboratory data will be summarized separately:

- Cycle 1 results (from all patients in the safety population) up to the day a patient starts Cycle 2 or 10 weeks after starting Cycle 1, whichever comes first
- Cycle 2 results (from patients in the safety population who receive at least 2 cycles of study drug) up to the day a patient starts Cycle 3 or 10 weeks after starting Cycle 2, whichever comes first
- Cycle 3 results (from patients in the safety population who receive at least 3 cycles of study drug) up to the day a patient starts Cycle 4 or 10 weeks after starting Cycle 3, whichever comes first
- Cycle 4 results (from patients in the safety population who receive 4 cycles of study drug) up to 10 weeks after starting Cycle 4

Laboratory test results measured on a continuous scale and changes from baseline values will be summarized within treatment arm using mean, standard deviation, median, minimum and maximum values. Results will be summarized at baseline, on Days 1-9, and every 7 days starting with Day 15 for patients receiving ACM, and at baseline, on Days 1-4, and every 7 days starting with Day 8 for patients receiving CM. Up to the point of starting a new treatment cycle or withdrawing from the study, a patient's most recent test result will be carried forward and included in analyses on days when clinical laboratory testing was not performed.

Ordinal categorical test results will be summarized within treatment arm using shift tables. Additionally, for tests where CTCAE [version 4.03] severity criteria are specified, CTCAE severity grades will be summarized in shift tables.

7.11.5. Vital Signs

Vital signs and changes from baseline values will be summarized by visit and within treatment arm using mean, standard deviation, median, minimum and maximum values. Results will be summarized at baseline, on Day 1 of each dosing cycle, and every 7 days thereafter. A patient's most recent test result will be carried forward and included in analyses on days when vital signs were not assessed. Typical comparisons of results summarized by visit may have limited interpretability for the same reasons outlined in section 7.11.4. Therefore, vital signs will be analyzed separately for the four groups identified there.

7.11.6. Concomitant Medication Use

Optionally, depending on the perceived relevance, concomitant medications will mapped to terminology in the WHO DDE and then summarized within treatment arm by usage rates for each level-1 ATC term and preferred (i.e., standardized) drug name.

7.12. Subgroup Analyses

8. Adverse Event and Medication Coding Dictionaries

The MedDRA and WHO DD versions (editions) that will be used to standardize medical terminology are not known at this time, but both will be recent editions published within 2 years before the date the first patient is randomized into the study and the versions used will be documented in the clinical study report.

9. Summary Tables

Suggested titles for summary tables to be included in the clinical study report are shown in Table 9 and blueprints for the format and content of these tables are presented in Appendix 2.

Table 9. Planned Summary Tables

Summary Table [*comments*]	Patient Population		
	ITT	Safety	Per-Protocol
Patient Disposition	X		
Demography	X	X	X
AML History	X	X	X
Medical History	X	X	X
Pre-Existing Medical Conditions	X	X	X
Relevant Prestudy Medications	X	X	X
Study Drug Exposure	X	X	X
Complete Remission Rates – Independent Hematologists Review [*repeat as needed for subgroup analyses*]	X		X
Complete Remission Rates – Local Investigator Review [*repeat as needed for subgroup analyses*]	X		X
Combined CR and Combined Remission Rates – Independent Hematologists Review [*repeat as needed for subgroup analyses*]	X		X
Combined CR and Combined Remission Rates – Local Investigator Review [*repeat as needed for subgroup analyses*]	X		X
Correlation Between Independent Hematologists and Local Investigator Reviews – Complete Remission	X		X
Correlation Between Independent Hematologists and Local Investigator Reviews – All Levels of Remission	X		X
Progression-Free Survival [*repeat as needed for subgroup analyses*]	X		X
Progression-Free Survival Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Overall Survival [*repeat as needed for subgroup analyses*]	X		X
Overall Survival Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X

Summary Table [*comments*]	Patient Population		
	ITT	Safety	Per-Protocol
Relapse-Free Survival Following CR [*repeat as needed for subgroup analyses*]	X		X
Relapse-Free Survival Following CR Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Relapse-Free Survival Following Combined CR [*repeat as needed for subgroup analyses*]	X		X
Relapse-Free Survival Following Combined CR Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Relapse-Free Survival Following Combined Remission [*repeat as needed for subgroup analyses*]	X		X
Relapse-Free Survival Following Combined Remission Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Complete Remission Duration [*repeat as needed for subgroup analyses*]	X		X
Complete Remission Duration Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Combined Complete Remission Duration [*repeat as needed for subgroup analyses*]	X		X
Combined Complete Remission Duration Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Combined Remission Duration [*repeat as needed for subgroup analyses*]	X		X
Combined Remission Duration Censoring at Transplant [*repeat as needed for subgroup analyses*]	X		X
Stem Cell Transplant Rates	X		X
ECOG Performance Status Shift Table	X		X
30- and 60-Day Mortality		X	
Numbers of Patients with Adverse Events		X	
Treatment-Emergent Adverse Events: Incidence Rates by MedDRA System Organ Class		X	
Treatment-Emergent Adverse Events: Incidence Rates by MedDRA Preferred Term		X	
Treatment-Emergent Adverse Events Judged Related to Study Drug: Incidence Rates by MedDRA System Organ Class		X	
Treatment-Emergent Adverse Events Judged Related to Study Drug: Incidence Rates by MedDRA Preferred Term		X	

Summary Table [*comments*]	Patient Population		
	ITT	Safety	Per-Protocol
Treatment-Emergent Adverse Events in the Standardized MedDRA Query for Tumor Lysis Syndrome: Incidence Rates by MedDRA Preferred Term		X	
Treatment-Emergent Adverse Events in the Standardized MedDRA Query for Tumor Lysis Syndrome Judged Related to Study Drug: Incidence Rates by MedDRA Preferred Term		X	
Serious Adverse Events: Incidence Rates by MedDRA System Organ Class		X	
Serious Adverse Events: Incidence Rates by MedDRA Preferred Term		X	
Serious Adverse Events Judged Related to Study Drug: Incidence Rates by MedDRA System Organ Class		X	
Serious Adverse Events Judged Related to Study Drug: Incidence Rates by MedDRA Preferred Term		X	
Treatment-Emergent Adverse Events Leading to a Delay in the Administration of Study Drug: Incidence Rates by MedDRA Preferred Term		X	
Treatment-Emergent Adverse Events Leading to a Reduction in the Protocol-Specified Dose of Study Drug: Incidence Rates by MedDRA Preferred Term		X	
Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug: Incidence Rates by MedDRA Preferred Term		X	
Treatment-Emergent Adverse Events Leading to Withdrawal from the Study: Incidence Rates by MedDRA Preferred Term		X	
Fatal Adverse Events: Incidence Rates by MedDRA Preferred Term		X	
Tumor Lysis Syndrome Incidence Rates During Cycle 1 by Maximum Cairo-Bishop Severity Grade		X	
Tumor Lysis Syndrome Incidence Rates During Study by Maximum Cairo-Bishop Severity Grade		X	
Clinical Laboratory Test Results and Changes from Baseline		X	
Clinical Laboratory Test Result Shift Tables		X	
Vital Signs and Changes from Baseline		X	
Concomitant Medications		X	

10. References

1. O'Brien PC, Fleming TR. A Multiple Testing Procedure for Clinical Trials. *Biometrics*. 1979;35:549-556.
2. Grimwade D, Walker H, Oliver F, et al. The Importance of Diagnostic Cytogenetics on Outcome in AML: Analysis of 1,612 Patients Entered Into the MRC AML 10 Trial. *Blood*. Oct 1 1998;92:2322-2333.
3. Yue L. Testing Multiple Secondary Endpoints in Confirmatory Comparative Studies – A Regulatory Perspective. Presented at: 2nd Annual AdvaMed Statistical Issues for Medical Devices and Diagnostics; April 29, 2009; Bethesda, MD. Available at: http://www.advamedmtli.org/download/File/Statistics_April_2009/Yue_lilly_pres.pdf. Accessed February 5, 2015.
4. Koch GG, Gansky SA. Statistical Considerations for Multiplicity in Confirmatory Protocols. *Drug Information J*. 1996;30:523-533. Available at: http://ferran.torres.name/download/shared/general/Multiplicity%20dj302612.pdf. Accessed February 5, 2015.
5. Mosteller RD. Simplified Calculation of Body-Surface Area. *N Engl J Med*. 1987;317:1098
6. Mantel N, Haenszel W. Statistical Aspects of Analysis of Data from Retrospective Studies of Disease. *J Natl Cancer Inst*. 1959;22:719-748.
7. Mantel N. Chi-Square Tests with One Degree of Freedom: Extensions of the Mantel-Haenszel Procedure. *J Am Stat Assoc*. 1963;58:690-700.
8. SAS Institute Inc. 2010. Base SAS® 9.2 Procedures Guide: Statistical Procedures, Third Edition. Cary, NC: SAS Institute Inc.
9. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas*. 1960;20:37-46.
10. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc*. 1958;53:457-481.
11. Pierceall WE, Lena R, Medeiros BC, et al. Mcl-1 Dependence Predicts Response to Vorinostat and Gemtuzumab Ozogamicin in Acute Myeloid Leukemia. *Leuk Res*. 2014;38:564-568.
12. Efron B, Stein C. The Jackknife Estimate of Variance. *Annals Stat*. 1981;9:586-596.
13. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. U.S. Dept of Health and Human Services, National Institutes of Health, National Cancer Institute; May 28, 2009 (v4.03 June 14, 2010). Available at: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf. Accessed February 16, 2015.
14. Cairo MS, Bishop M. Tumor Lysis Syndrome: New Therapeutic Strategies and Classification. *Br J Haematol*. 2004;127:3-11.

Appendix 1. Criteria for Grading Severity of Clinical Tumor Lysis Syndrome

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Renal	Creatinine ≤1.5 x ULN	Creatinine >1.5-3.0 x ULN	Creatinine >3.0-6.0 x ULN	Creatinine >6.0 x ULN	Death
Cardiac Arrhythmia	Intervention not indicated	Non-urgent medical intervention indicated	Symptomatic and incompletely controlled medically or controlled with device (e.g., defibrillator)	Life-threatening (e.g., arrhythmia associated with CHF, hypotension, syncope, shock)	Death
Seizure	—	One brief generalized seizure; seizure(s) well controlled by anticonvulsants or infrequent focal motor seizures not interfering with activities of daily living	Seizure in which consciousness is altered; poorly controlled seizure disorder; with breakthrough generalized seizures despite medical intervention	Seizures of any kind which are prolonged, repetitive or difficult to control (e.g., status epilepticus, intractable epilepsy)	Death

Clinical TLS requires one or more clinical manifestations (renal, cardiac, neuro) along with criteria for laboratory TLS
Maximal clinical manifestation defines the severity grade
ULN = upper limit of normal; 114.4 and 105.6 µmol/L will be used for males and females, respectively, if ULN is not available for an institution

Appendix 2. Sample Summary Table Templates

Blueprint for summaries of continuous data within and between treatment arms (delete p-value if summary is for safety data; delete fourth column if summary is only within treatment arm):

Study TPI-ALV-201 | Title for Summary Table (<ITT | Safety | Per-Protocol> Patient Population) | Page x of x

	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
Characteristic 1			
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
Median	xx.x	xx.x	xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		
Characteristic 2			
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
Median	xx.x	xx.x	xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		
Characteristic 3			
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
Median	xx.x	xx.x	xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		

[a] p-value from the Student *t* test

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of time-to-event data within and between treatment arms (delete fourth column if summary is only within treatment arm):

Study TPI-ALV-201 | Title for Summary Table (<ITT | Safety | Per-Protocol> Patient Population) | Page x of x

	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
Characteristic 1			
Median	xx.x	xx.x	xx.x
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
25%, 75% Percentiles	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		
Characteristic 2			
Median	xx.x	xx.x	xx.x
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
25%, 75% Percentiles	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		
Characteristic 3			
Median	xx.x	xx.x	xx.x
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
25%, 75% Percentiles	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		

[a] p-value from the log-rank test

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of categorical data within and between treatment arms (delete p-value if summary is for safety data; delete fourth column if summary is only within treatment arm):

Study TPI-ALV-201 | Title for Summary Table (<ITT | Safety | Per-Protocol> Patient Population) | Page x of x

	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
Characteristic 1			
Category A	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Category B	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Category C	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
⋮	⋮	⋮	⋮
p-value[a]	0.xxx		
Characteristic 2			
Category A	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Category B	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Category C	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
⋮	⋮	⋮	⋮
p-value[a]	0.xxx		

[a] p-value from the Fisher exact test

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Summaries of continuous, time-to-event and categorical data may be combined similar to this blueprint for AML history (delete fourth column if summary is only within treatment arm):

Study TPI-ALV-201 | AML History (<ITT | Per-Protocol> Patient Population) | Page x of x

	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
Months Since AML Dx			
Median	xx.x	xx.x	xx.x
95% CI	(xx.x, xx.x)	(xx.x, xx.x)	(xx.x, xx.x)
25%, 75% Percentiles	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
p-value[a]	0.xxx		
Clinical Onset of AML			
De Novo AML	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Prior MDS	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Prior Leukemogenic Tx	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)	(xx.x%, xx.x%)
p-value[b]	0.xxx		
⋮	⋮	⋮	⋮

[a] p-value from the log-rank test
[b] p-value from the Fisher exact test

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of CR rates:

Study TPI-ALV-201 Complete Remission Rates Based on <Central | Local> Hematologist's Review (<ITT | Per-Protocol> Patient Population) Page 1 of 1

	ACM (n=xxx)	CM (n=xxx)
Complete Remission Rates		
Number (%) w/o Bone Marrow[a]	xx (xx.x%)	xx (xx.x%)
CR Rate (95% CI)	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
p-value[b]	0.xxx	
Odds Ratio[c] (95% CI)	xx.x (xx.x, xx.x)	

[a] patients rated as not achieving CR due to not providing bone marrow biopsy/aspirate
[b] p-value from CMH general association test stratified by [*list stratification factors used for randomization*]
[c] Mantel-Haenszel odds ratio stratified by [*list stratification factors used for randomization*]

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of CRi, CRp, PR, combined CR and combined remission rates:

Study TPI-ALV-201 | CRi, CRp, PR, Combined CR and Combined Remission Rates Based on <Central | Local> Hematologist's Review (<ITT | Per-Protocol> Patient Population) | Page x of x

	ACM (n=xxx)	CM (n=xxx)
CRi Rates		
Number (%) w/o Bone Marrow[a]	xx (xx.x%)	xx (xx.x%)
CRi Rate (95% CI)	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
p-value[b]	0.xxx	
Odds Ratio[c] (95% CI)	xx.x (xx.x, xx.x)	
CRp Rates		
Number (%) w/o Bone Marrow[a]	xx (xx.x%)	xx (xx.x%)
CRp Rate (95% CI)	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
p-value[b]	0.xxx	
Odds Ratio[c] (95% CI)	xx.x (xx.x, xx.x)	
PR Rates		
Number (%) w/o Bone Marrow[a]	xx (xx.x%)	xx (xx.x%)
PR Rate (95% CI)	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
p-value[b]	0.xxx	
Odds Ratio[c] (95% CI)	xx.x (xx.x, xx.x)	
⋮	⋮	⋮

[a] patients rated as not achieving response due to not providing bone marrow biopsy/aspirate
[b] p-value from CMH general association test stratified by [*list stratification factors used for randomization*]
[c] Mantel-Haenszel odds ratio stratified by [*list stratification factors used for randomization*]

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of correlation between central and local hematologist reviews (delete rows and columns for CRi, CRp and PR if summary is CR assessments only):

Study TPI-ALV-201 | Correlation Between Central and Local Hematologist Reviews <Complete | All Levels of> Remission (ITT Patient Population[a]) | Page 1 of 1

	Local Review					
Central Review	CR	CRi	CRp	PR	NR[b]	Total
CR	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
CRi	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
CRp	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
PR	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
NR[b]	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Total	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (100%)

Kappa (correlation) coefficient[c]: K=0.xxx

[a] Table summarizes all post-baseline slides from ITT population, therefore a patient may be counted more than once
[b] NR = no response
[c] x slides only reviewed by the central hematologist and x slides only reviewed by the local investigator (not tabled) are not included in the calculation of the kappa coefficient

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of time-to-event efficacy endpoints (delete rows for Treatment Failure and Relapse After CR as appropriate):

Study TPI-ALV-201 | Title for Summary Table (<ITT | Per-Protocol> Patient Population) | Page 1 of 1

	ACM (n=xxx)	CM (n=xxx)
<PFS | OS | RFS | Resp. Duration>		
Treatment Failure (%)	xx (xx.x%)	xx (xx.x%)
Relapse After CR (%)	xx (xx.x%)	xx (xx.x%)
Death (%)	xx (xx.x%)	xx (xx.x%)
Censored (%)	xx (xx.x%)	xx (xx.x%)
First Quartile (95% CI)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
Median (95% CI)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
Third Quartile (95% CI)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
Minimum, Maximum	xx.x, xx.x	xx.x, xx.x
Survival Estimates (95% CI)		
6 months	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
12 months	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
18 months	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
p-value[a]	0.xxx	
Hazards Ratio[b] (95% CI)	x.xx (x.xx, x.xx)	
Hazards Ratio[c] (95% CI)	x.xx (x.xx, x.xx)	

[a] footnote origin of p-value
[b] based on Cox proportional hazards model with treatment arm as the only independent variable
[c] based on Cox PH model with effects for treatment arm and [*list stratification factors used for randomization*]

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of stem cell transplant rates:

Study TPI-ALV-201 | Stem Cell Transplant Rates (<ITT | Per-Protocol> Patient Population) | Page 1 of 1

	ACM (n=xxx)	CM (n=xxx)
Stem Cell Transplant Within 90 Days Following Last Dose of Study Drug		
SCT Status Unknown[a]	xx (xx.x%)	xx (xx.x%)
SCT Rate (95% CI)	xx.x% (xx.x%, xx.x%)	xx.x% (xx.x%, xx.x%)
p-value[b]	0.xxx	
Odds Ratio[c] (95% CI)	xx.x (xx.x, xx.x)	

[a] patients lost to follow-up (without documentation of SCT) within 90 days following last dose of study drug; these patients are counted as not receiving SCT
[b] p-value from CMH general association test stratified by [*list stratification factors used for randomization*]
[c] Mantel-Haenszel odds ratio stratified by [*list stratification factors used for randomization*]

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for shift tables:

Study TPI-ALV-201 | Title for Summary Table | Page x of x
Treatment Arm: ACM | (<ITT | Safety | Per-Protocol> Patient Population)

	Worst <Score | Grade> During Study						
Baseline <Score | Grade>	0	1	2	3	4	5	Total
0	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
1	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
2	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
3	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
4	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
5	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Total	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

Treatment Arm: CM

	Worst <Score | Grade> During Study						
Baseline <Score | Grade>	0	1	2	3	4	5	Total
0	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
1	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
2	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
3	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
4	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
5	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Total	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for summaries of relevant prestudy and concomitant medications within and between treatment arms:

Study TPI-ALV-201	<Relevant Prestudy | Concomitant> Medications (<ITT | Safety | Per-Protocol> Patient Population)		Page x of x
WHO Drug Dictionary Enhanced Anatomical and Therapeutic Class	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
ATC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 1.1 under ATC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 1.2 under ATC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 1.3 under ATC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 1.4 under ATC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
⋮	⋮	⋮	⋮
ATC 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 2.1 under ATC 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 2.2 under ATC 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 2.3 under ATC 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Drug 2.4 under ATC 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
⋮	⋮	⋮	⋮

A patient is counted once if he or she received any <relevant prestudy | concomitant> medications coded to the WHO Drug Dictionary preferred drug name.

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx; WHO DDE [version]

Blueprint for summaries of relevant pre-existing medical conditions and adverse events within and between treatment arms (delete SOC for summaries sorted by preferred term):

Study TPI-ALV-201 | Title for Summary Table (<ITT | Safety | Per-Protocol> Patient Population) | Page x of x

MedDRA SOC MedDRA Preferred Term	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
SOC 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
PT 1.1 under SOC 1			
Any Grade	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 4	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 3	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
⋮	⋮	⋮	⋮
PT 2.1 under SOC 1			
Any Grade	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 4	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 3	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
⋮	⋮	⋮	⋮

A patient is counted once if he or she experienced any event coded to the MedDRA term. The highest reported severity grade is presented.

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx; MedDRA v#

Blueprint for summary of 30- and 60-day mortality

Study TPI-ALV-201	30- and 60-Day Mortality (Safety Patient Population)	Page 1 of 1
	ACM (n=xxx)	CM (n=xxx)
Number of Deaths	xx	xx
30-Day Mortality[a]	xx.x%	xx.x%
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)
Number of Deaths	xx	xx
60-Day Mortality[a]	xx.x%	xx.x%
95% CI	(xx.x%, xx.x%)	(xx.x%, xx.x%)

[a] derived from Kaplan-Meier curves for Overall Survival

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for high-level safety summary (include additional page to show [1] TEAE leading to a reduction in the protocol-specified dose of study drug, [2] TEAE leading to discontinuation of study drug, and [3] TEAE leading to withdrawal from the study)

Study TPI-ALV-201 | Numbers of Patients with Adverse Events (Safety Patient Population) | Page 1 of 2

Number of Patients with any…	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
TEAE			
Grade 5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 4 or Higher	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 3 or Higher	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 2 or Higher	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Any Grade	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Treatment-Related TEAE			
Grade 5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 4 or Higher	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 3 or Higher	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 2 or Higher	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Any Grade	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Treatment-Emergent SAE	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
TEAE Leading to a Delay in the Administration of Study Drug	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

TEAE = Treatment-Emergent Adverse Event; SAE = Serious Adverse Event

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx

Blueprint for TLS incidence rates by Cairo-Bishop severity grade

Study TPI-ALV-201 Tumor Lysis Syndrome Incidence Rates During [Cycle 1 | Study] by Maximum Cairo-Bishop Severity Grade (Safety Patient Population) Page 1 of 1

Number of patients with…	ACM (n=xxx)	CM (n=xxx)	All Patients (n=xxx)
No TLS	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Laboratory TLS[a]	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Clinical TLS			
Grade 1	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 2	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 3	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 4	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Grade 5	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)

[a] Without clinical manifestations

SAS v9.4 (dd-mmm-yyyy hh:mm) from ADxx, ADxx, ADxx