STATISTICAL ANALYSIS PLAN

A Multicenter, Randomized, Double-blind, Placebo-controlled Phase 2 Trial to Evaluate the Efficacy and Safety of OCV-501 in Elderly Patients With Acute Myeloid Leukemia

NCT Number: NCT01961882

PRT NO.: 311-12-001

Version Date: 09 January 2018 (Version 3.0)

PAREXEL International

Otsuka Pharmaceutical Co., Ltd.

Protocol No. 311-12-001

A Multicenter, Randomized, Double-blind, Placebo-controlled Phase 2 Trial to Evaluate the Efficacy and Safety of OCV-501 in Elderly Patients With Acute Myeloid Leukemia

Statistical Analysis Plan

PAREXEL Project Number: 210876

Version 3.0
Date: 09 Jan 2018

Table of Contents

LIST OF ABBREVIATIONS

Abbreviation	Expansion
AE	Adverse event
ADR	Adverse drug reaction
AML	Acute myeloid leukemia
AR(1)	First-order autoregressive
ATC	Anatomic Therapeutic Chemical Classification
CI	Confidence interval
CS	Compound symmetry
CSH	Heterogeneous CS
CTCAE	Common Terminology Criteria for Adverse Events
DFS	Disease-free survival
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EORTC QLQ-C30	The European Organization for Research and Treatment of Cancer QLQ-C30
FAS	Full analysis set
HF	Harrington-Fleming
IFN-γ	Interferon gamma
HR	Hazard ratio
IRS	Immune response set
IMP	Investigational medicinal product
IVRS	Interactive Voice Response Services
IWRS	Interactive Web Response Services
LS	Least squares
MedDRA	Medical Dictionary for Regulatory Activities (The ICH international medical dictionary)
MMRM	Mixed-effect model repeated measures
OS	Overall survival
PPS	Per protocol set
PS	Performance status
PT	Preferred term
QTc	Corrected QT interval
QTcF	QT interval as corrected by Fridericia's formula
QOL	Quality of life
REML	Restricted maximum likelihood
RS	Raw score
SAE	Serious adverse event
SD	Standard deviation
SOC	System organ class
SS	Safety analysis set
TOEP	Toeplitz

Abbreviation	Expansion
UN	Unstructured
VC	Variance Components
WHO-DD	World Health Organization Drug Dictionary
WT1	Wilm's tumor gene 1

1 INTRODUCTION

OCV-501 is a drug product containing an HLA class II restricted Wilm's tumor gene 1 (WT1) derived peptide.

In the phase 1 trial (Trial 311-10-001), OCV-501 was subcutaneously administered once a week for 4 times at doses of 0.3 mg, 1 mg, and 3 mg in elderly acute myeloid leukemia (AML) patients and safety and tolerability of OCV-501 were assessed. In this phase 1 trial, no dose-limiting toxicities were reported and the tolerability of OCV-501 treatment up to a dose of 3 mg was confirmed. In addition, OCV-501 specific delayed-type hypersensitivity was observed at each dose. This finding suggests that OCV-501 induces the T cell mediated immune response.

This trial is planned as a randomized placebo-controlled comparative trial, to assess the efficacy and safety of OCV-501 treatment as a maintenance therapy in elderly patients with AML who have completed consolidation therapy after being judged to have achieved complete remission following an induction regimen.

This statistical analysis plan is based upon the following study documents:

- Study protocol,
- Electronic case report form (eCRF)

2 TRIAL OBJECTIVES

The primary objective is to compare the disease-free survival (DFS) in AML patients 60 years or older, who are randomly assigned to receive either OCV-501 monotherapy or placebo.

The secondary objectives are:

- To compare overall survival (OS) between both treatment arms
- To compare patient quality of life (QOL) (European Organization for Research and Treatment of Cancer [EORTC] QLQ-C30) and performance status (PS; Eastern Cooperative Oncology Group performance status [ECOG PS]) between both treatment arms
- To characterize and compare safety in both treatment arms
- To assess immunological responses (OCV-501 specific interferon gamma [IFN-γ] production, WT1-killer peptide specific IFN-γ production, anti-OCV-501 antibody level, anti-WT1 antibody level, immunoglobulin)
- To assess WT1 mRNA level

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This trial is a multinational, multicenter, randomized, double-blind, placebo-controlled trial in elderly patients with AML, to compare DFS in patients 60 years or older with AML who are randomly assigned to receive either OCV-501 monotherapy or placebo.

This trial consists of the following three periods.

3.1.1 Screening Period

The eligible patients will be randomly assigned to one of the following treatment arms at a ratio of 1:1

- OCV-501 arm: 3 mg of OCV-501 (0.4 mL)
- Placebo arm: Placebo (0.4 mL)

The patients will be assigned to either treatment arm via the interactive voice response services (IVRS) or the interactive web response services (IWRS).

3.1.2 Treatment Period

From Week 1 to Week 104, the investigational medicinal product (IMP) will be subcutaneously administered: once-weekly up to the 8th administration, and every two weeks from the 9th administration onward.

Blood sampling for expressed level of WT1mRNA, anti-OCV-501 antibody, anti-WT1 antibody, and immunoglobulin will be performed at least once every four weeks.

OCV-501 specific IFN-γ production, WT1-killer peptide specific IFN-γ production will be assessed at Week 5, Week 9, Week 13, and every 12 weeks.

QOL and PS will be measured at least once every 4 weeks. A concomitant drug survey will be performed at every scheduled visit. For the safety of the subjects, a medical interview, blood biochemical tests, urinalysis, and adverse event (AE) observation will be performed. Vital signs, body weight, and 12-lead electrocardiogram (ECG) will be measured.

Retrospective evaluation of samples will be conducted by the central pathological review committee. The central pathological review committee will evaluate the sample used at the judgment of relapse by the trial site and the previous sample from before relapse, and verify the judgment of relapse of the trial site.

3.1.3 Post-treatment Observation Period

Subjects will return to the trial site 11 to17 days after the last dose of the 2-year IMP administration period to undergo a completion examination during the post-treatment observation period. If a patient discontinues IMP administration before completion of the 2-year treatment period, the patient will undergo a withdrawal examination within 7 days after the investigator or sub-investigator judged withdrawal to be necessary.

The schedule of observations, examinations, and evaluations is shown in Table 3.1.3-1.

Table 3.1.3-1 Trial Schedule

Procedure (Week) (Day)	Screening period	Treatment period															Post-treatment observation period *1	
		W1 to W8 (once-weekly IMP administration)								W9 to W104 (IMP administration every two weeks)							**Completion**	**Withdrawal**
		W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W11**	**W13**	**W15**	After Week 17 up to Week 102	**W103**	**W104**	11 - 17 days after W 103	Within 7 days from withdrawal judgment
		D1	D8	D15	D22	D29	D36	D43	D50	D57	D71	D85	D99		D715	D722	D279	
(Allowance)		1	5 - 11	12 - 18	19 - 25	26 - 32	33 - 39	40 - 46	47 - 53	54 - 60	66 - 76	80 - 90	94 - 104	Same as W11 to W15 (up to Week 102)	710 - 720	-	11-17 days after W 103	
Written informed consent	●																	
Subject Information	□																	
Virus test [a)]	□																	
Chest x-ray	□																	
hCG pregnancy test [b)]	□																	
Body height	□																	
Patient enrollment for this trial		○																
Blood sampling for HLA genotyping		○																
IMP administration [c)]		↓ (1)	↓ (2)	↓ (3)	↓ (4)	↓ (5)	↓ (6)	↓ (7)	↓ (8)	↓ (9)	↓ (10)	↓ (11)	↓ (12)	↓ (every 2 weeks) (13-55)	↓ (56)			
Concomitant drug survey	←	←	←	←	←	←	←	←	←	←	□	→	→	→	→	→	→	→
Adverse event observation	←	←	←	←	←	←	←	←	←	←	□	→	→	→	→	→	→	→
Medical interview by (sub) investigator	□	○	○	○	○	○	○	○	○	○	○	○	○	○ (every 2 weeks)	○		□	□
Body weight	□	▲				□				□		□		□ (every 4 weeks)			□	□
Vital signs [d)]	□	▲				○				○		○		○ (every 4 weeks)			□	□
Hematology test [e) f)]	□	▲		○		○		○		○	○	○	○	○ (every 2 weeks)	○		□	□
Peripheral blood smear preparation	□	▲		□		□		□		□	□	□	□	□ (every 2 weeks)	□		□	□
Blood biochemistry test [g)]	□	▲				○				○		○		○ (every 4 weeks)			□	□
12-Lead electrocardiogram	□					○				○		○					□	□
QOL (EORTC QLQ-C30)		○				□				□		□		□ (every 4 weeks)			□	□
Performance status (ECOG PS)	□	▲				□				□		□		□ (every 4 weeks)			□	□
Bone marrow aspiration [f)]	□																□	
Bone marrow smear preparation [g)]	□																□	
Urinalysis [h)]	□	▲				○				○		○		○ (every 4 weeks)			□	□
Expressed level of WT1mRNA		○				○				○		○		○ (every 4 weeks)			□	□
IFN-γ production (OCV-501 specific, WT1-killer peptide specific)		○				○				○		○		○ (at Weeks 25, 37, 49, 61, 73, 85, 97)			□	□
Anti-501 antibody		○				○				○		○					□	□
Anti-WT1 antibody		○				○				○		○		○ (every 4 weeks)			□	□
Immunoglobulin		○				○				○		○					□	□

●: Obtained before screening examinations.

○: Mandatory (to be performed/collected before IMP administration).

▲: Data from a screening examination can be used if they were obtained within 7 days before the IMP administration (to be performed before IMP administration except peripheral blood smear preparation).

□: Mandatory

[a] HIV antibody, HBV-DNA, HCV antibody.

[b] To be performed only for female subjects who are capable of becoming pregnant (Urine test).

[c] Once-weekly up to the 8th administration, once every two weeks from the 9th administration onward. The subject's general conditions, including injection site reaction, must be observed at the trial site for one hour after IMP administration.

[d] Blood pressure, pulse, body temperature

[e] Hb, Hct, RBC, WBC, WBC differential (visual count), PLT.

[f] If leukemic blast cells appear in peripheral blood or if there is a reduction in blood cell components, perform bone marrow aspiration by post-treatment observation period, and prepare smear.

[g] AST, ALT, ALP, LDH, γ-GTP, TP, ALB, T-Bil, BUN,UA, Cr

[h] pH and specific gravity, bilirubin, glucose, ketone bodies, white blood cells, occult blood, protein, urobilinogen

* [1] Even if treatment is discontinued before completion of the 2-year treatment period, patients will still undergo the post-treatment examination.
(If an AE has not resolved by the end of the post-treatment observation period, the investigator or subinvestigator will explain to the subject the need for follow-up investigation and will request the subject's cooperation. The investigator or subinvestigator will conduct a follow-up investigation within 4 weeks after the post-treatment observation period.)

3.2 Efficacy and Safety Variables

3.2.1 Efficacy

Primary endpoint:

- Disease-free survival (DFS)

Secondary endpoint:

- Overall survival (OS)
- Quality of life (EORTC QLQ- C30)
- Performance status (ECOG PS)

Reference endpoint:

- DFS until OS cutoff date

3.2.2 Safety

AE, hematology tests, blood biochemistry tests, urinalysis, vital signs (blood pressure, pulse rate, body temperature), body weight and ECG

3.2.3 Exploratory variables

OCV-501 specific IFN-γ production, WT1-killer peptide specific IFN-γ production, anti-OCV-501 antibody level, anti-WT1 antibody level, immunoglobulin and WT1 mRNA level

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency and integrity in accordance with standard PAREXEL procedures.

4.2 General Presentation Considerations

The important terms will be defined as follows:

Term	Definition
Screen failure	A screen failure is a subject from whom written informed consent was obtained, but to whom an IMP was not allocated
Individual subject trial start date	The day of obtaining the subject's written informed consent
Individual subject withdrawal	A subject who discontinue OCV-501 or placebo treatment before completion of the post-treatment observation period
Individual subject trial withdrawal date	The day of the withdrawal examination of the post-treatment observation period within 7 days after the investigator or subinvestigator judged withdrawal to be necessary
Individual subject completion	A subject who complete 2 years of treatment with OCV-501 or placebo, and complete the post-treatment examination at 11-17 days after the end of IMP administration
Individual subject trial completion date	The day of the completion examination at 11-17 days after the end of the 2-year IMP administration period
Individual subject trial period	The period from the day of obtaining the subject's informed consent to the day of trial completion or the day of the withdrawal examination. Does not include the follow-up period nor the outcome survey.

Continuous data will be summarized in terms of the mean, standard deviation (SD), median, minimum, maximum and number of observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point, frequency counts and percentages. Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will be calculated using the number of observations with non-missing values as the denominator.

Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. However, if a p-value is only presented to four decimal places (by SAS) it will not be rounded again, but will be presented to four decimal places. P-values less than 0.001 will be presented as "<0.001".

Two sided statistical tests with α=0.05 will be performed to test the hypotheses.

95% Confidence intervals (CIs) will be presented to one more decimal place than the raw data. Multiplicity will not be adjusted.

All report outputs will be produced using SAS® version 9.3 or a later version in a secure and validated environment.

Baseline

Baseline is defined as the last available screening or pre-treatment assessment.

Allowance window

Assessments taken outside of protocol allowable windows will be included in the summaries according to the eCRF assessment recorded by the investigator.

Unscheduled assessments

Extra assessments (laboratory data, vital signs or exploratory items associated with non-protocol clinical visits or obtained in the course of investigating or managing AEs) will be included in listings but not summaries.

Discontinuation assessments

Discontinuation assessments will be summarized according to the eCRF assessments recorded by the investigator at the time of discontinuation, not otherwise specified.

Graphs

Graphs will be plotted by "day" or "month" or "Visit timepoint (week)" as specified in the shells.

4.3 Study Subjects

4.3.1 Disposition of Subjects

A clear accounting of the disposition of all subjects who enter the study from screening to study completion will be generated.

The following summaries will be generated.

- A summary of subjects with informed consents, screen failure subjects by reasons, subjects randomized, entering and discontinuing each period (week 1-8, week 9-104 and post-treatment), subjects completed the study and discontinued subjects (Analysis population: all subjects with informed consents)
- A summary of the number of subjects randomized per site and per country/region (Analysis population: all subjects randomized)
- A summary of subjects treated (with at least one IMP administration), subjects completed the study and discontinuations from the study with major reasons. (Analysis population: all subjects randomized)

4.3.2 Protocol Deviations

The protocol deviation summaries may include the following:

- A summary of subjects with major protocol deviation in each country/region and site by treatment arm and overall and by type of deviation (Analysis population: all subjects randomized)
- A summary of the number and percentage of subjects with major deviations (Inclusion/Exclusion Criteria, Withdrawal Criteria, IP administration/Study treatment, Disallowed Medication)
- A by-subject listing of major or minor protocol deviations will be generated.

4.4 Analysis Populations

In this study, 4 analysis populations are employed, and the definitions are described in the following section. The analysis population summaries include the following:

- A summary of the number and percentage of subjects by treatment arm for each analysis population will be generated.
- A by-subject listing of analysis population details (site, subject ID inclusion/exclusion flag for each population and reason for exclusion from each population) will also be generated.

4.4.1 Safety Analysis Set

The safety analysis set (SS) includes all subjects who have received the IMP at least once and from whom data on at least one safety endpoint was obtained after the start of IMP administration.

If subjects have received the wrong IMP administration other than the one that they were originally randomized into, then they will be analyzed as "As treated".

4.4.2 Full Analysis Set

The full analysis set (FAS) includes all subjects who received the IMP at least once and from whom data on at least one efficacy endpoint was obtained after the start of IMP administration.

If the subjects receive the wrong IMP administration, then they will be analyzed as "As randomized".

4.4.3 Per Protocol Set

The per protocol set (PPS) includes all FAS subjects except those who have major protocol deviations or those who have withdrawn from the trial within 8 weeks from start of IMP administration. However, if a subject withdrew due to any reason other than an event (relapse or death) within 8 weeks from start of IMP administration, the subject will still be included in the PPS.

Subjects with the following major protocol deviations are excluded from PPS.

- Inclusion/Exclusion Criteria
- Compliance deviations (Patients do not receive the IMP at least 2 times up to the 8th administration or compliance rate is less than 80% throughout the period.)
- Disallowed medications
- Withdrawal Criteria

If the subjects receive the wrong IMP administration, then they will be analyzed as "As randomized".

4.4.4 Immune Response Set

The immune response set (IRS) includes all subjects who received the IMP at least once and from whom data on at least one immune response endpoint were obtained after the start of IMP administration.

If the subjects receive the wrong IMP administration, then they will be analyzed as "As treated".

4.5 Demographic and Other Baseline Characteristics

1) Definition:

Demographic:

Age, sex, height, body weight (at screen), race, trial country/region

Age will be divided into 3 groups (60 to 64 years, 65 to 70 years, ≥ 71 years).

Underlying diseases, induction/consolidation therapy and other baseline factors

Diagnosis of AML (WHO classification), duration from the last dose of consolidation therapy to randomization, duration from diagnosis of AML to randomization, regimen of number and therapy (induction, consolidation), WT1 mRNA level at Day 1, unfavorable karyotype abnormality (abn(3q)/inv(3)/t(3;3) or -5/del(5q) or -7/del(7q) or t(6;9)), cytogenetic abnormality (FLT3-ITD mutation)

WT1 mRNA level (copy/ugRNA) will be divided into 4 categories (<50, 50-199, 200<= and Unknown).

Prior and concomitant drugs/treatments

Prior and concomitant medications will be classified by Anatomic Therapeutic Chemical Classification 2 (ATC2) and generic (preferred) name using World Health Organization Drug Dictionary (WHO-DD).
The start or stop dates of administration of drugs/treatments other than IMP will be compared to the date of first IMP administration to classify them as either prior or concomitant drugs/treatments.

Table 4.5-1 Prior and Concomitant Drugs/Treatments

Prior	Stop date of drug/treatment ≤ First IMP administration date
Concomitant	Stop date of drug/treatment > First IMP administration date

Concomitant drugs/treatments starting after the completion/withdrawal date will be listed but will not be classified or summarized.

Medical history and Concomitant disease

Medical history and concomitant disease are coded using Medical Dictionary for Regulatory Activities (MedDRA version 20.1). If the condition is not continuing at the

time of informed consent, it is considered as medical history. If the condition is continuing, then it is concomitant disease.

2) Analysis Set:

FAS, PPS, SS, IRS

3) Analysis Method:

Frequency counts and percentage or descriptive statistics for each variable will be calculated according to the nature of the data (continuous or discrete) by treatment arm and overall. For WT1 mRNA level, the measurement "<50" will be translated into 49 when calculating descriptive statistics.

The following summaries will be generated.

- A summary of demographic variables at baseline by treatment arm and over all
- A summary of other baseline characteristics by treatment arm and overall
- Subgroup analysis (Age, Country/Region) of demographic characteristics (FAS only)
- Subgroup analysis (Age, Country/Region) of other baseline characteristics (FAS only)
- A summary of prior and concomitant drugs/treatments (WHO drug terms) by treatment arm and overall (FAS only)
- A summary of Medical history by System Organ Class (SOC) and Preferred Term (PT) by treatment arm and overall (FAS only)
- A summary of Concomitant disease by SOC and PT by treatment arm and overall (FAS only)

4.6 Treatment Compliance

1) Definition:

- Treatment compliance will be calculated as

$$Treatmentcompliance(\%) = \frac{Actual\ number\ of\ IMP}{Scheduled\ number\ of\ IMP} \times 100,$$

  Where the scheduled number of IMP administrations is the number scheduled until the time of withdrawal.

2) Analysis Set:

SS

3) Analysis Method:

A summary of the descriptive statistics will be calculated by treatment arm. Compliance rates (%) will be divided into 6 categories (<20, 20 - <40, 40 - <60, 60 - <80, 80 - <100, 100). The following outputs will be generated.

- A summary of IMP exposure by compliance rate
- A by-subject listing of treatment compliance data

4.7 Efficacy Evaluation

FAS will be used as the primary analysis population and PPS will be used as the secondary analysis population to evaluate the stability.

4.7.1 Primary Efficacy Endpoints

4.7.1.1 Disease-free Survival

1) Definition:

- Observation period: For each subject, the time from randomization to the DFS-cutoff date. The DFS-cutoff date was set as the completion/discontinuation date of each subject.
- DFS: The time from randomization until relapse or death from any cause, whichever comes first by the DFS-cutoff date.
- Event: Relapse or death from any cause by the DFS-cutoff date.
- Censoring:
    - If an event has not been observed by the DFS-cutoff date, then the last confirmed non-relapse date will be used as the censoring date.
    - For subjects who have withdrawn from this trial before the completion date for reasons other than relapses or deaths, if non-relapse has been confirmed on the withdrawal date, then the withdrawal date will be used as the censoring date. Otherwise, the last confirmed non-relapse date will be used as the censoring date.

2) Analysis Set:

FAS (primary), PPS

For the judgment of relapse by the central pathological review committee, DFS will be evaluated in the FAS and PPS as primary analysis.

For the judgment of relapse by trial site, DFS will only be evaluated in the FAS as secondary analysis.

3) Analysis Method:

- Primary test for comparison between OCV-501 and Placebo will be conducted by Log-rank test. For exploratory purposes, the Harrington-Fleming (HF) test will be performed by applying the weight 1- $\hat{S}(t)$, where $\hat{S}(t)$ is the product-limit estimates of survival functions.
- Kaplan-Meier estimate of survivor function $\hat{S}(t)$ and curve

  $\hat{S}(t)$ = Pr (Survive beyond t)

$$= \prod_{j:t_j<t} \left(\frac{n_j - y_j}{n_j} \right)$$

  where t_j is the j-th event time, n_j is the number of patients at risk and y_j is the number of events.

The Kaplan-Meier curve plots the estimated survival function $\hat{S}(t)$ vs. time separately for each treatment arm.

- The 2-year survival rate and median survival time will be calculated. The 95% CI of the 2-year survival rate will be calculated using Greenwood formula.
- Hazard ratio (HR) and its 95% CI will be calculated from the Cox proportional hazards model.
  - In the model of Cox analysis, treatments effect will be adjusted with covariates (Age, Country/Region and WT1 mRNA level at Day 1). The model with covariates (WT1 mRNA only) or model without covariates will be also analyzed. In these models, age will be categorized into 3 groups (60 to 64 years, 65 to 70 years, ≥ 71 years), Country/Region into 2 groups (Japan, Republic of Korea+Taiwan), WT1 mRNA into 4 groups (<50, 50-199, 200<= and Unknown).
  - BRESLOW method will be used for ties.
  - Proportionality assumption in Cox proportional hazard model will be assessed using 2 methods:
    1) SAS PHREG procedure ASSESS PH option will be used for checking proportional hazard assumption.
    2) Look at the log-minus-log survival plots graphically. If the graph results in parallel lines then the proportionality assumption is satisfied.
- HR and its 95% CI at every 3 months interval will be calculated by adding a time varying covariate to Cox Proportional Hazards model as an alternative analysis if the proportionality assumption fails. The model will be as follows:

  $\lambda(t;X) = \lambda_0(t) \exp(\beta_1 trt*period\_1 + \beta_2 trt*period\_2 + ...)$

  where $\lambda_0(t)$ is the baseline hazard function,

trt is the treatment arm(1= OCV-501 and 0= Placebo) and

$$period\_1 = \begin{cases} 1 & if\ 0 \le DFS\ time < 3\ Months \\ 0 & otherwise \end{cases}$$

$$period\_2 = \begin{cases} 1 & if\ 3 \le DFS\ time < 6\ Months \\ 0 & otherwise \end{cases}$$

...

So HR (OCV-501 vs. Placebo) for each period will be calculated as follows:

λ(OCV-501|period_k)= λ_0(t)exp(β_k)

λ(Placebo|period_k)= λ_0(t)

HR(OCV-501 vs. placebo|period_k)= λ_0(t)exp(β_k)/ λ_0(t)= exp(β_k)

95% CI :[exp(β_k-1.96* $\sqrt{Var(\beta_k)}$), exp(β_k+1.96* $\sqrt{Var(\beta_k)}$)]

- For exploratory purposes, the incidence rate over the entire follow-up period will be calculated using the person-year method by treatment arm as follows.

  1) Calculate the number of events and person-year at risk by treatment arm.

  2) Calculate the incidence rates and its corresponding 95% CI by treatment arm. (1 year is calculated as 365.25 days)

  Rate= $\frac{Number\ of\ events}{Person-year}$

  95% CI:

  [Rate/exp(1.96* $\sqrt{\frac{1}{Number\ of\ events}}$), Rate* exp(1.96* $\sqrt{\frac{1}{Number\ of\ events}}$)]

4.7.2 Secondary Efficacy Endpoints

4.7.2.1 Overall Survival

1) Definition:

– Observation period: For all subjects, the time from randomization to OS-cutoff date. The OS-cutoff date was set as the date after 728 days (2 years) from the day that the last subject started IMP administration without consideration of withdrawal by the OS-cutoff date.

- Definition: The time from randomization until death from any cause by the OS-cutoff date.
- Event: Death from any cause by the OS-cutoff date.
- Censoring: If an event has not been observed by the OS-cutoff date, then the last confirmed survival date will be used as the censoring date.

2) Analysis Set:

FAS, PPS

3) Analysis Method:

The same analysis will be conducted as for DFS.

4.7.2.2 Quality of Life

1) Based on the scoring rule (EORTC QLQ-C30 version 3), the following scales and single items will be calculated and analyzed.

Table 4.7.2.2-1 Quality of Life Scale
Global health status / QOL
Functional Scale: Physical, Role, Cognitive, Emotional, Social
Symptom Scale: Fatigue, Pain, Nausea/Vomiting
Single Items: Dyspnea, Sleep disturbance, Appetite loss, Constipation, Diarrhea, Financial impact

Scoring rule[13]:

The QLQ-C30 is composed of both multi-item scales and single-item measures. These include 5 functional scales, 3 symptom scales, a global health status/QOL scale, and 6 single items. Each of the multi-item scales includes a different set of items - no item occurs in more than one scale.

- All of the scales and single-item measures range in score from 0 to 100.

- A high scale score represents a higher response level. Thus a high score for a functional scale represents a high/healthy level of functioning, a high score for the global health status/QOL represents a high QOL, but a high score for a symptom scale/item represents a high level of symptomatology/problems.

Scoring principle:

1) Estimate the average of the items that contribute to the scale; this is the raw score (RS).

2) Use a linear transformation to standardize the RS, so that scores range from 0 to 100; a higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms

Technical summary:

In practical terms, if items I_1, I_2,.... , I_n are included in a scale, the procedure is as follows:

Raw score (RS)

Calculate the RS:

$$\text{Raw score} = RS = \frac{I_1 + I_2 + \Lambda\ I_n}{n}$$

Linear transformation

Apply the linear transformation to 0-100 to obtain the score S

Functional scales: $S = \left\{1 - \frac{(RS - 1)}{Range}\right\} \times 100$

Symptom scales / items: $S = \left\{\frac{(RS - 1)}{Range}\right\} \times 100$

Global health status / QOL: $S = \left\{\frac{(RS - 1)}{Range}\right\} \times 100$

- Range is the difference between the maximum possible value of RS and the minimum possible value. The QLQ-C30 has been designed so that all items in any scale take the same range of values. Therefore, the range of RS equals the range of the item values. Most items are scored 1 to 4, giving range = 3.
- The exceptions are the items contributing to the global health status/QOL, which are 7-point questions with range = 6, and the initial yes/no items on the earlier versions of the QLQ-C30 which have range = 1.

2) Analysis Set:

FAS, PPS

3) Analysis Method:

- A summary of descriptive statistics of each score will be calculated by treatment arm.

- Mixed-effect model repeated measures (MMRM) will be performed with treatment as fixed effect.
- The dependent variable is the change from baseline to each visit in each scale of QOL score. The model will include treatment arm, protocol-specified visit, and treatment-by-visit interaction.
- Changes from baseline at each time point will be plotted in trend diagrams. The difference of Least square (LS)-means and CI between two treatment arms will also be generated.

- The primary hypothesis test is the test of the difference of LS-means between treatment arms over time.

- The mixed model is as follows:

$$Y = X\beta + Zu + \varepsilon$$

- A key assumption is that u and ε are normally distributed with

$$E\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and Var} \begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

- The mean will be modeled as E[Y]= Xβ and the variance will be modeled as Var [Y] = ZGZ' + R. Therefore, the variance of Y can be modeled by setting up the structure of G and R.

- The covariance structure will be assumed to be UN. It could be switched from UN to TOEP (Toeplitz), CSH (Heterogeneous Compound Symmetry), AR(1) [(Autoregressive(1)], CS (Compound Symmetry) and VC (Variance Components) in this order if convergence is not achieved.

- Restricted maximum likelihood (REML) will be used to make the inference.

- All analyses will be performed on the observed data.

- Discontinuation assessment will be analyzed using the following allowance window. If two values fall into the same visit window, then the latter will be taken into analysis.

Week	Target Day[a]	Study Day Interval[a]		
1(Baseline)	1	-7	-	1
5	29	2	-	43
9	57	44	-	71
13	85	72	-	99
17	113	100	-	127
21	141	128	-	155
25	169	156	-	183
29	197	184	-	211

33	225	212	-	239
37	253	240	-	267
41	281	268	-	295
45	309	296	-	323
49	337	324	-	351
53	365	352	-	379
57	393	380	-	407
61	421	408	-	435
65	449	436	-	463
69	477	464	-	491
73	505	492	-	519
77	533	520	-	547
81	561	548	-	575
85	589	576	-	603
89	617	604	-	631
93	645	632	-	659
97	673	660	-	687
101	701	688	-	715
105	729	716	-	743

a. Relative to the first day of IMP taken in treatment phase (Date of first IMP taken is defined as the Day 1)

4.7.2.3 Performance Status

1) Definition: Performance status (PS) will be scored according to the following criteria.

Table 4.7.2.3-1 Performance Status Grade

Grade	ECOG
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g, light house work, office work
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
5	Dead

2) Analysis Set:

FAS, PPS

3) Analysis Method:

- Frequency distributions of PS scores at the baseline and each post-baseline time point (including completion/discontinuation) will be calculated by treatment arm.

- Shift table of PS scores by grade at baseline and the end of trial will be generated by treatment arm.
- PS will be compared at the time of completion or discontinuation based on the improvement factor, which will be defined as follows.

Table 4.7.2.3-2 Improvement Factor

Improvement factor	Definition
Better	Current PS score < Baseline PS score
No change	Current PS score = Baseline PS score
Worse	Current PS score > Baseline PS score

- To compare the treatment effect in PS improvement factor between treatment arms, Wilcoxon 2-sample test will be carried out.

4.7.3 Reference Efficacy Endpoint

4.7.3.1 DFS until OS cutoff date

1) Definition:

- Observation period: For all subjects, the time from randomization to OS cutoff date
- DFS for reference: The time from randomization until relapse or death from any cause, whichever comes first, by the OS cutoff date
- Event:
  Relapse or death from any cause, whichever comes first by the OS cutoff date
- Censoring:
  - If an event has not been observed by the OS cutoff date, the last confirmed non-relapse date will be used as the censoring date.
  - For subjects who have withdrawn from this trial before the OS cutoff date for reasons other than relapse or death, if non-relapse has been confirmed on the withdrawal date, then the withdrawal date will be used as the censoring date. Otherwise, the last confirmed non-relapse date will be used as the censoring date.

2) Analysis Set:

For the judgment of relapse by trial site, DFS will only be evaluated in the FAS.

3) Analysis Method:

The same analysis method will be conducted as for DFS.

4.7.3.2 Therapeutic Response

1) Definition:

Relapse will be assessed according to the response criteria of the International Working Group shown in the protocol, Table 7.2.1.1-1. Relapse will be judged by the investigator in the site (Site judgment) and by Central Pathological Reviewer (Central judgment) and recorded in eCRF.

2) Analysis Set:

FAS, PPS

3) Analysis Method:

- A summary of the number and percentage of subjects judged as Relapse by both Central judgment and Site judgment will be calculated.

4.8 Safety Evaluation

4.8.1 Extent of Exposure

1) Definition:

Duration of exposure to treatment (days), full dose of OCV-501, total number of doses

2) Analysis Set:

SS

3) Analysis Method:

A summary of the descriptive statistics will be calculated by treatment arm.

Duration of exposure to treatment will be calculated in days as follows:

IMP administration end date − IMP administration start date + 1

4.8.2 Adverse Events

1) Definition:

AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA version 20.1). Severity will be graded using Common Terminology Criteria for Adverse Events (CTCAE) version 4.03.

Treatment-emergent AEs (TEAE) are defined as those AEs that either start or worsen in severity on or after the date/time of first dose of IMP administration. TEAEs will be tabulated by SOC and PT with PT substituted for verbatim terms.

An adverse drug reaction (ADR) is defined as a response to a drug which is noxious and unintended, and which occurs at doses normally used in man for the prophylaxis, diagnosis, or therapy of disease, or for the modifications of physiological function (WHO definition). In this study, AEs with causal relationship (Definitely related, Probably related, Possibly related) are considered as ADR (potentially drug related AEs).

When dates are missing or partially missing, AEs will be assumed to be treatment-emergent, unless there is clear evidence to suggest that the AE started prior to the first dose of IMP administration.

Also, serious adverse event (SAE), death and AE leading to discontinuation will be considered separately.

2) Analysis Set:

SS

3) Analysis Method:

All TEAE summaries will describe the number of subjects reporting at least one AE and the total number of events reported. These will generally include the following summaries:

- A summary of the number and percentage of subjects reporting a TEAE or ADR by treatment arm, SOC and PT
- A summary of the most common TEAEs or ADRs by treatment arm, SOC and PT (reported by ≥5% of subjects in any treatment arm)
- A summary of the TEAEs by treatment arm and PT (reported by ≥5% of subjects in any treatment arm)
- A summary of the number and percentage of subjects reporting a TEAE or ADR by treatment arm, CTCAE severity grade, SOC and PT
- A summary of the number and percentage of subjects reporting a TEAE by maximum CTCAE grade and PT (Category: All grade, Grade 1+2, Grade 3+4+5)
- A summary of the number and percentage of subjects reporting a TEAE or ADR leading to death by treatment arm, SOC and PT
- A summary of the number and percentage of subjects reporting a serious TEAE by treatment arm, SOC and PT

- A summary of the number and percentage of subjects reporting a TEAE or ADR leading to discontinuation by treatment arm, SOC and PT
- A summary of all the summaries above

For by-severity summaries, for each subject reporting multiple AEs, the worst severity recorded will be used.

By-subject listings of all AEs (including non-treatment-emergent events) will be generated. These will include the following.

- A by-subject listing of all deaths that occurred during the study
- A by-subject listing of all SAEs other than death
- A by-subject listing of all AEs leading to discontinuation of IMP administration

4.8.3 Clinical Laboratory Evaluation

1) Definition:

A subject will be defined as having a treatment-emergent laboratory abnormality if any of the following conditions are satisfied for a specific laboratory parameter.

- A laboratory result within the normal range at baseline and either a result below the lower limit of the normal range or above the upper limit of the normal range at any post-baseline time point.
- A laboratory result below the lower limit of the normal range at baseline and a laboratory result above the upper limit of the normal range at any post-baseline time point.
- A laboratory result above the upper limit of the normal range at baseline and a laboratory result below the lower limit of the normal range at any post-baseline time point.

In addition, laboratory values will be graded using CTCAE version 4.03.

2) Analysis Set:

SS

3) Analysis Method:

The following summaries will be generated:

- A summary of each laboratory parameter and its change from baseline by treatment arm and time point

- Shift table of laboratory parameter (Urinalysis) at each post-baseline point
- A summary of the number and percentage of subjects experiencing low, normal and high values at baseline and at each post-baseline time point (shift table)
- A summary of the number and percentage of subjects with different CTCAE grades at baseline, at each post-baseline time point and worst grade in post-baseline (shift table)
- A summary of new or worsened laboratory abnormalities based on CTCAE Grades (Categories: all grades, Grade 1+Grade 2, Grade 3+Grade 4)
  In this table, the worst grade observed in post-baseline are summarised. Patients with no change or better grade from baseline are not counted. 'New' abnormality means 'grade 0' at baseline and '>= grade 1' after baseline. All grade means '>= grade 1' after baseline.

By-subject listings of all laboratory data will be generated by treatment arm with reference range provided and abnormal values will be marked with flags (H/L, Gx).

4.8.4 Vital Signs, Physical Findings and Other Observations Related to Safety

1) Definition:

Vital signs (blood pressure, pulse rate, body temperature), body weight and ECG.

ECG parameters will include heart rate, PR interval, QRS complex, QT interval and QTc interval.

2) Analysis Set:

SS

3) Analysis Method [13]:

The following summaries will be generated:

- A summary of each vital signs, body weight and ECG and its change from baseline by treatment arm and time point.
- A summary of the number and percentage of subjects having QTcF interval values > 450 ms, > 480 ms and > 500 ms at each time point.
- A summary of the number and percentage of subjects having QTcF interval changes from baseline > 30 ms and > 60 ms at each post-baseline time point.
- Shift table of 12-Lead ECG findings at each post-baseline time point and worst in post-baseline assessment

By-subject listings of vital sign parameters and ECG results will also be generated.

4.9 Safety Interim Analysis

For the purpose of data safety monitoring, the safety interim analysis will be conducted regularly on all subjects while the level of blinding is being maintained.

The following summaries will be generated.

- A summary of patient background
- A summary of treatment compliance (overall and until withdrawal)
- A summary of number and percentage of subjects reporting a TEAE by SOC and PT
- A summary of number and percentage of subjects reporting a TEAE by CTCAE severity grade, SOC and PT

The following by-subject listings will be generated.

- A by-subject listing of all SAEs other than death
- A by-subject listing of all deaths

Also, the following graphs will be generated.

- Kaplan-Meier curve of DFS
- Kaplan-Meier curve of OS

4.10 Exploratory Items Analysis

1) Definition:

OCV-501 specific IFN-γ production, WT1-killer peptide specific IFN-γ production, anti-OCV-501 antibody level, anti-WT1 antibody level, WT1 mRNA level and immunoglobulin

2) Analysis Set:

IRS

3) Analysis Method:

Generally, a summary will be calculated in terms of the mean, geometric mean, SD, coefficient of variation, median, minimum, maximum and number of patients, unless otherwise stated.

At each time point, a summary of each exploratory parameter and change from baseline will be generated and time courses of immune response values will be plotted by treatment arms.

4.11 Subgroup Analysis

1) Definition:

DFS, OS, safety (Summary of TEAE, TEAE by MedDRA SOC and PT) and exploratory parameters will be examined for the following subgroups.

Table 4.11-1 Subgroup

Factor	Categories
Country/ Region	Japan, Republic of Korea, and Taiwan
Age	60 to 64 years, 65 to 70 years, ≥ 71 years

2) Analysis Set:

FAS for DFS and OS, SS for safety, IRS for Immune response

3) Analysis Method:

For each subgroup, the following statistical analyses will be performed.

DFS/OS: log-rank test, HF test, KM estimates of DFS/OS rate
Safety: Summary of TEAE, Incidences of TEAE by MedDRA SOC and PT
Exploratory parameters: Summary of descriptive statistics

4.12 Determination of Sample Size

The number of AML patients is not very high, and is estimated to be approximately 7000 in Japan[7]. From the fact that in the previous clinical trial (JALSG GML200) in elderly AML patients, it has been forecast that the possible enrollment in Japan is to be at most 70 in one year. Also, from the preliminary trial feasibility survey, it is known that around 50 subjects could be recruited in one year in the Republic of Korea and Taiwan. Through a global trial in Japan, the Republic of Korea and Taiwan, it is considered possible to enroll 120 subjects in one year to assess the efficacy of OCV-501.

Within the limitation of the past experiences with regard to IMP administration, the target sample size will be considered with reference to following publications.[5,6,7,9] With reference to disease-free survival which has been reported from trials in AML patients 60 years or older who have received the consolidation therapy followed by maintenance therapy, the 2-year DFS rate of placebo arm has been estimated to be around 20%.[5,6,7]

On the other hand, considering the so-called graft-versus-leukemia (GVL) effect associated with a lower relapse risk after the allogeneic hematopoietic stem cell transplantation, OCV-501 is basically expected to have the same benefits. With reference to the findings of Kurosawa et al,[9] which have reported a 30% higher 2 or 3-year DFS in elderly AML subjects who have received the allogeneic hematopoietic stem cell transplantation compared with those without transplantation, OCV-501 is expected to have a 15% - 20% higher 2-year DFS. Under the assumption of exponential distribution of DFS time, the hazard ratio (placebo vs. OCV-501) has been set to 0.57 - 0.65, and with a two-sided log-rank test at the 5% significance level, 60 or more subjects in each treatment group can achieve a power of 53% - 75% or more to detect the difference between placebo and OCV-501 groups.

4.13 Changes in the Conduct of the Study or Planned Analysis

Any changes made after the database lock will be recorded in the relevant section of the clinical study report.

Changes from the analysis planned in the protocol are as follows:

Section number	Contents of changes	Reason of changes
4.5 Demographics and Other Basic Characteristics	WT1 mRNA level at Day 1 was added instead of AML diagnosis or at the time of induction therapy or consolidation therapy.	Day 1 data would be more influential than the data at diagnosis or at the time of induction or consolidation therapy.
4.7.1.1 DFS 4.7.2.1 OS	Proportionality assumption will be tested first. Cox PH model will be conducted on every 3 month interval as an alternative analysis.	If proportionality assumption fails, then Cox PH model might be considered invalid to conduct on the whole period.

Section number	Contents of changes	Reason of changes
4.7.1.1 DFS 4.7.2.1 OS	Age, country and other factor (WT1 mRNA) will be adjusted as covariates in the Cox PH model. Furthermore, subgroup analysis will be conducted.	Cox PH model analyses adjusted by covariates are added to investigate the influence of covariates. And the subgroup analyses are added as supplementary analyses.
4.7.2.2 QOL	Mixed-effect model repeated measures will be conducted.	To make a comparison on QOL between two treatment arms.
4.7.2.3 PS	Improvement factor will be created and Wilcoxon 2-sample test will be conducted.	To make a comparison on PS between two treatment arms.
4.7.3. Reference efficacy endpoint 4.7.3.1 DFS until OS cutoff date 4.7.3.2 Therapeutic Response	New efficacy variables have been added.	For the purpose of reference
4.8.2 Adverse Events	A summary of SAEs by severity is not conducted.	Listing will cover it.
4.10 Exploratory Items Analysis	For WT1 mRNA item, t-test is not conducted.	According to the nature of WT1mRNA data, comparison between treatment arms will not be necessary.
4.11 Subgroup analysis	Safety and immune response have been added to the subgroup analysis.	For the purpose of reference

5 REFERENCES

1 Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, U.S. Department of Health and Human Services Food and Drug Administration , May 2007, Clinical/Medical

2 Little, R.C, et al. 2006. SAS for Mixed Model, Second Edition, Cary, NC: SAS Institute Inc.

3 SAS Online Doc 9.1.3. http://support.sas.com/onlinedoc/913/docMainpage.jsp

4 Kincaid C. 2005. Guidelines for selecting the covariance structure in mixed model analysis. SUGI 30 Proceedings: Paper 198-30.

5 Baer MR, George SL, Caligiuri MA, Sanford BL, Bothun SM, Mrózek K, et al. Low-dose interleukin-2 immunotherapy does not improve outcome of patients age 60 years and older with acute myeloid leukemia in first complete remission: Cancer and Leukemia Group B Study 9720. J Clin Oncol. 2008, 26(30):4934-9.

6 Hellstrand K, Romero A, Brune M. Immunotherapy with Histamine Dihydrochloride and Interleukin-2 in Acute Myeloid Leukaemia. European Haematology, 2008;2(1):71-4

7 Farag SS, George SL, Lee EJ, Baer M, Dodge RK, Becknell B, et al. Postremission therapy with low-dose interleukin 2 with or without intermediate pulse dose interleukin 2 therapy is well tolerated in elderly patients with acute myeloid leukemia: Cancer and Leukemia Group B study 9420. Clin Cancer Res. 2002;8(9):2812-9.

8 Portal Site of Official Statics of Japan (e-Stat): Statistics Bureau, Ministry of Internal Affairs and Communications and National Statistics Center: from: http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do

9 Kurosawa S. Comparison of post-remission strategies in elderly patients with acute myeloid leukemia in first complete remission, Rinsho Ketsueki. 2011;52(8):645-52.

10 Klaus Rostgaard Methods for stratification of person-time and events- a prerequisite for Poisson regression and SIR estimation: Epidemiologic Perspectives & Innovations 2008, 5:7

11 Leslie E. Daly Confidence limits Made easy: Interval Estimation using a Substitution: American Journal of Epidemiology, Vol 147, No 8

12 EORTC QLQ C30 version 3.0

13 EORTC QLQ C30 Scoring Manual Third edition, 2001

14 ICH Harmonized Guideline, THE CLINICAL EVALUATION OF QT/QTC INTERVAL PROLONGATION AND PROARRHYTHMIC POTENTIAL FOR NON-ANTIARRHYTHMIC DRUGS (E14), 2005-05-12