STATISTICAL ANALYSIS PLAN

ATLAS: A Phase 2, Open-label Study of Rucaparib in Patients with Locally Advanced or Metastatic Urothelial Carcinoma

PROTOCOL NUMBER:	CO-338-085
VERSION:	1
DATE FINAL:	31 January 2018
SPONSOR:	Clovis Oncology, Inc. [redacted]

Approval Page

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND SPECIALIST TERMS

AE(s)	adverse event(s)
ACT	anti-cancer therapy
ADP	adenosine diphosphate
ATC	Anatomical Therapeutical Chemical (coding)
BID	twice daily
BRCA	breast cancer gene
BRCA1	breast cancer gene 1
BRCA2	breast cancer gene 2
CI	confidence interval
C_{min}	trough plasma concentration
CR	complete response
CrCL	creatinine clearance
CRF	Case Report Form
CTCAE	Common Terminology Criteria for Adverse Events
ct DNA	circulating cell-free tumor DNA
CV	coefficient of variation
DMC	data monitoring committee
DOR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	End of Treatment
FMI	Foundation Medicine, Inc.
HRD	homologous recombination deficiency
HRR	homologous recombination repair
ICH	International Council on Harmonisation
invPFS	progression-free survival, investigator-assessed
ITT	intent-to-treat
LOH	loss of heterozygosity
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NGS	next-generation sequencing
ORR	objective response rate

OS	overall survival
PALB2	partner and localizer of BRCA2
PARP	poly(ADP-ribose) polymerase
PD	progressive disease
PFS	progression-free survival
PK	pharmacokinetic(s)
PR	partial response
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SI	International System of Units
SNP(s)	single-nucleotide polymorphism(s)
StD	standard deviation
TEAE	treatment-emergent adverse event
WHO	World Health Organization

1 INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analyses and data summaries to be performed to assess the efficacy, safety, and pharmacokinetics (PK) of rucaparib (CO-338) for Clovis Oncology, Inc sponsored clinical study CO-338-085, entitled "ATLAS: A Phase 2, Open-label Study of Rucaparib in Patients with Locally Advanced or Metastatic Urothelial Carcinoma."

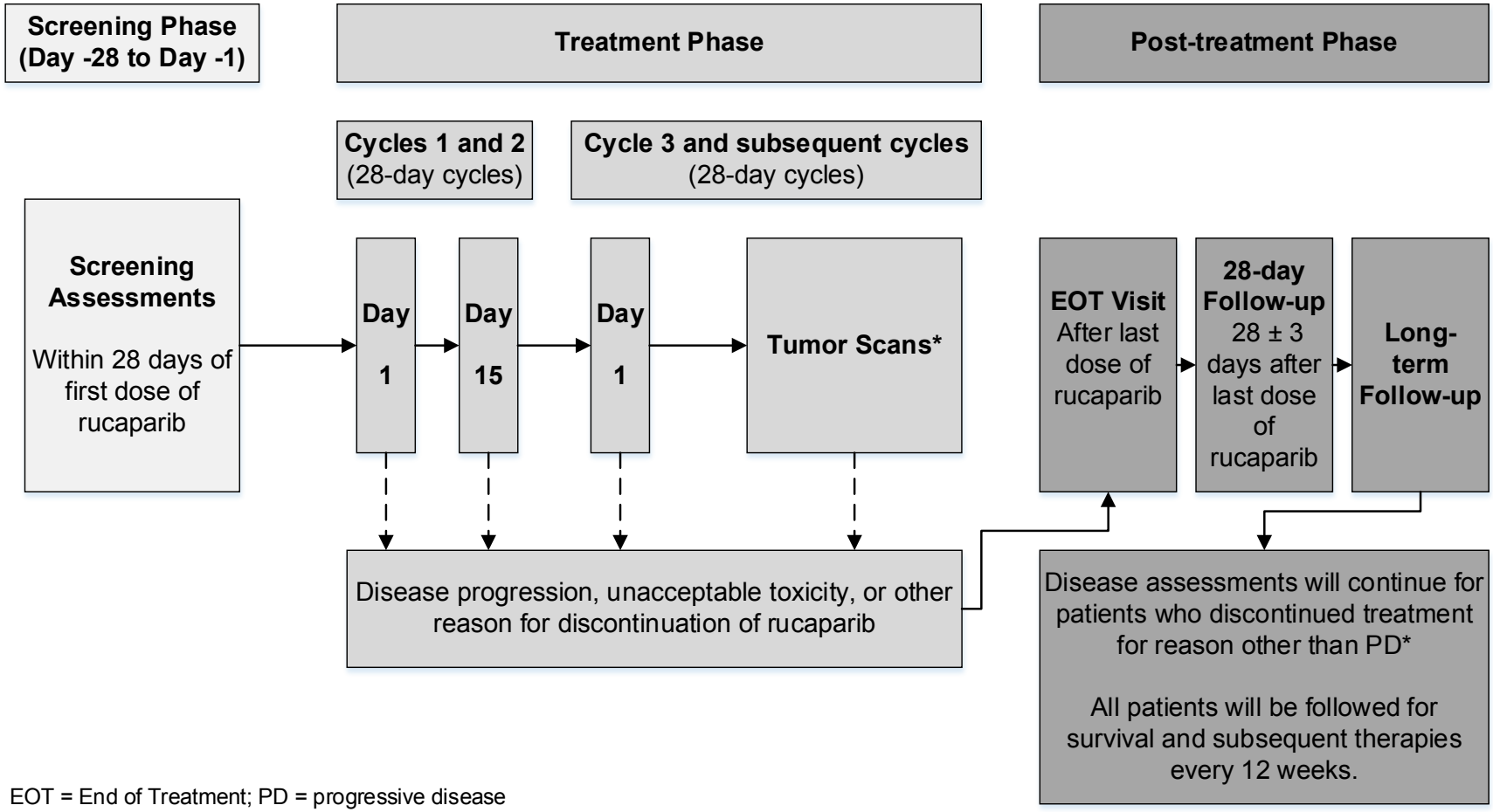
2 OVERALL STUDY DESIGN, OBJECTIVES, AND ENDPOINTS

2.1 Study Design

This clinical trial is a Phase 2 multicenter, open-label study evaluating rucaparib 600 mg twice daily (BID) for treatment of patients with recurrent locally advanced or metastatic urothelial carcinoma. This study will enroll patients with measurable disease per Response Evaluation Criteria in Solid Tumors (RECIST) Version 1.1 (v1.1). All patients must have received 1 or 2 prior standard of care treatment regimens and have radiologic progression during or after the most recent regimen. Patients who have received prior poly(adenosine diphosphate [ADP]-ribose) polymerase (PARP) inhibitor treatment will be excluded. Tumor tissue collected prior to rucaparib treatment will be analyzed to determine homologous recombination deficiency (HRD) status. This HRD status will be utilized to classify patients into molecularly-defined subgroups for efficacy analyses. The criteria used to determine if a patient is HRD-positive has been prospectively defined in this SAP (see Section 3.2).

The study schema is shown in Figure 1. This study consists of a Screening Phase, a Treatment Phase, and a Post-treatment Phase. Patients will receive rucaparib monotherapy in the Treatment Phase, and will undergo procedures and assessments including regular safety and efficacy evaluations during the entire conduct of the study.

Figure 1. Study Schema



EOT = End of Treatment; PD = progressive disease
* Tumor scans every 8 calendar weeks (±7 days) from Cycle 1 Day 1 up to 18 months then every 12 calendar weeks (±7 days) thereafter

2.2 Study Objectives and Endpoints

Table 1. Primary, Secondary, and Exploratory Objectives and Endpoints

Primary Objectives	**Primary Endpoints**
To evaluate objective response rate (ORR) in molecularly-defined homologous recombination deficiency (HRD)-positive and intent-to-treat (ITT) populations using a prospectively defined molecular signature.	ORR per Response Evaluation Criteria in Solid Tumor (RECIST) Version 1.1 (v1.1), as assessed by investigator, in HRD-positive and ITT populations
Secondary Objectives	**Secondary Endpoints**
To evaluate duration of response (DOR)	DOR per RECIST v1.1, as assessed by investigator
To estimate progression-free survival (PFS)	Disease progression according to RECIST v1.1, as assessed by the investigator, or death due to any cause
To estimate overall survival (OS)	OS
To evaluate safety and tolerability of rucaparib	Incidence of adverse events (AEs), clinical laboratory abnormalities, changes in ECGs and vital signs, and dose modifications
To evaluate steady-state pharmacokinetics (PK) of rucaparib	Trough (C_{min}) level rucaparib concentrations
Exploratory Objectives	**Exploratory Endpoints**
To assess biomarkers that correlate with response to rucaparib	Clinical efficacy of rucaparib in biomarker defined cohorts
To evaluate circulating cell-free tumor DNA (ctDNA) as a molecular marker of response to rucaparib	Association of cancer-related mutations detected in baseline ctDNA samples with response to rucaparib
To assess molecular changes over time in plasma and tumor samples	Association of changes in cancer-related mutations detected in plasma and tissue samples over time with response to rucaparib
To evaluate biomarkers associated with resistance to rucaparib	Biomarkers associated with resistance to rucaparib

2.3 Sample Size Justification

The overall sample size was determined by considering the number of patients needed for adequate safety and activity assessment of HRD-positive patients. Based on the estimated

prevalence of 60% HRD-positive patients in this population, approximately 200 patients will be enrolled into the study. This would result in enrollment of approximately 120 HRD-positive patients. The overall sample size will enable robust estimates of clinical activity in the HRD-positive and ITT populations.

3 GENERAL ANALYSIS CONVENTIONS

Quantitative variables will typically be summarized using frequencies and percentages for appropriate categorizations and may also be summarized using descriptive statistics. For variables summarized with descriptive statistics, the following will be presented: N, mean, standard deviation (StD), median, minimum, and maximum. Categorical variables will be presented using frequencies and percentages.

The Kaplan-Meier methodology will be used to summarize time-to-event variables.[1,2] Estimates of the proportion surviving across time will be plotted, and if estimable, the 50th (median) percentile with the corresponding 95% confidence interval (CI) will be presented. The number of patients with events and the number of censored patients over time will also be displayed.

Baseline is defined as the last measurement on or prior to the first dose of rucaparib administration, unless otherwise specified.

All statistical analyses will be conducted with the SAS® System, Version 9.3 or higher.

3.1 Analysis Populations

The following analysis populations are defined for the study:

Safety Population – The safety population will include all patients who received at least 1 dose of rucaparib.

Intent-to-treat Population (ITT) – The ITT population will include all patients who received at least 1 dose of rucaparib.

HRD-positive Population – The HRD-positive population will include all patients who received at least 1 dose of rucaparib and are classified as HRD-positive by the prospectively-defined HRD signature (See Section 3.2).

3.2 Definition of HRD Signature

The HRD signature for this study is defined as the percentage of genome with loss of heterozygosity (LOH).[3] Patients whose tumors have genome-wide LOH ≥ 10% will be considered HRD-positive.

A mandatory tumor specimen is required for all patients; these tissue samples will be analyzed by Foundation Medicine, Inc. (FMI) using a next-generation sequencing (NGS) assay. This assay interrogates approximately 3,500 single-nucleotide polymorphisms (SNPs) across the genome and detects alterations in 310 genes.[4] The SNPs on this panel will be used

to quantify the percentage of the genome with LOH, a marker of HRD shown to be associated with PARP inhibitor sensitivity in patients with relapsed ovarian cancer. This study will assess if LOH is correlated with PARP inhibitor sensitivity in patients with recurrent bladder cancer.

4 PATIENT DISPOSITION

Patient disposition will be summarized using frequency counts and the corresponding percentages. The number of patients in each analysis population, number of patients discontinued, and the primary reason for treatment discontinuation will be summarized.

5 PROTOCOL DEVIATIONS

The number of patients with major protocol deviations (eg, inclusion or exclusion criteria) will be determined prior to data base lock and will be summarized with frequencies and percentages or provided in a patient listing.

Protocol deviations will not be used to exclude any patients from the efficacy analyses.

6 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

All demographic and baseline characteristics will be summarized for the safety, ITT, and HRD-positive populations.

6.1 Demographics

The demographic variables will be summarized with frequency tabulations and descriptive statistics. The demographic variables presented will include age, height, weight, gender, race, and region using the following categorizations:

- Age (years): ≤ 50, 51-60, 61-70, 71-80, 81-90, > 90;
- Height (cm): ≤ 75, > 75-100, > 100-125, > 125-150, > 150-175, > 175;
- Weight (kg): ≤ 50, > 50-75, > 75-100, > 100-125, > 125-150, > 150;
- Race: American Indian or Alaska Native, Asian, Black, Native Hawaiian or Pacific Islander, White, Other, Not Reported
- Region: North America, Europe/Other

6.2 Baseline Clinical Characteristics

The following variables will be summarized with frequency tabulations:

- Histology (pure urothelial and mixed)
- De novo metastases (Y/N)
- Eastern Cooperative Oncology Group (ECOG) performance status (0, 1)
- Time since diagnosis of primary tumor (months) (0 to 12, > 12 to 24, > 24)

- Hemoglobin concentration (< 10 g/dL and ≥ 10 g/dL)
- Creatinine Clearance (CrCL) (< 60 mL/min and ≥ 60 mL/min)
- Metastatic sites (visceral, liver, lymph node only)
- Number of prior anti-cancer regimens (1, 2, >2)
- Prior platinum chemotherapy (Y/N)
- Type of prior platinum chemotherapy (cisplatin-based, carboplatin-based)
- Prior immune checkpoint inhibitor therapy (Y/N)
- Prior intravesical BCG (Y/N)
- Prior intravesical chemotherapy (Y/N)
- Prior neoadjuvant chemotherapy (Y/N)
- Prior adjuvant chemotherapy (Y/N)
- Prior radiotherapy (Y/N)
- Smoking status (current, former, never)
- Previous cystectomy/nephrouretectomy: (Y/N)
- Number of Bellmunt risk factors (0, 1, 2, 3)

Descriptive statistics may also be used to summarize the continuous variables. Other baseline characteristic variables of interest may also be summarized if deemed appropriate.

6.3 Medical History

Medical history events will be classified using the Medical Dictionary for Regulatory Activities (MedDRA) classification system version 20.1 or more recent. Medical history data will be summarized using frequency tabulations by system organ class and preferred term.

7 STUDY DRUG EXPOSURE AND COMPLIANCE

The following variables will be summarized:

- Number of cycles initiated
- Duration of treatment (days, months)
- Dose intensity
- Number of dose reductions

Duration of treatment will be calculated as 1+ the number of days from study drug start date to date of study drug discontinuation. Dose intensity will be calculated as the actual dose received divided by the assigned dose amount. Descriptive statistics and frequencies/percentages for appropriate categorizations will be used to summarize study drug exposure variables.

8 PRIOR AND CONCOMITANT MEDICATIONS

All concomitant treatments documented during the study period will be summarized in frequency tabulations. Prior/concomitant medication coding will utilize World Health Organization (WHO) Drug version 2017MAR01DDE (Enhanced).

Separate data summaries of prior medications will be provided. Prior medications will be defined as those medications with both a start and a stop date that is before the day of the first dose of study drug administration. If either the start date and/or the stop date of the medication is missing so that it is unclear whether the medication was stopped prior to first dose of study drug administration then the medication will be included in the summary of the concomitant medications.

9 EFFICACY ANALYSIS

Efficacy analyses will be performed using the ITT and HRD-positive populations.

9.1 Primary Efficacy Analysis

The primary efficacy endpoint of ORR is defined as the proportion of patients with a confirmed response of complete response (CR) or partial response (PR) by RECIST v1.1 (Appendix 1) as assessed by the investigator. The ORR will be summarized with frequencies and percentages.

9.2 Secondary Efficacy Analyses

9.2.1 Duration of Confirmed Response

The duration of confirmed response is defined as the time from the date that a response is first documented per RECIST v1.1 by the investigator to the time that progression is first documented per RECIST v1.1 by the investigator, plus 1 day.

The duration of confirmed response will be analyzed and summarized in a Kaplan-Meier plot. Any patient with an ongoing response at the time of data cut-off for analysis will be censored at the date of the last post-baseline scan. In the case of a low censoring rate, the duration of response may be summarized with descriptive statistics.

9.2.2 Progression-free Survival

Progression-free survival (PFS) is defined as the time from first dose of rucaparib to the date of first objective evidence of progression documented by the investigator per RECIST v1.1 or death due to any cause, whichever occurs first, plus 1 day. Patients without a documented event of progression will be censored on the date of their last adequate post-baseline tumor scan or date of first dose of study drug if no tumor scans have been performed. If a patient discontinues study drug due to something other than PD and subsequently begins another anti-cancer therapy (ACT) prior to a PFS, that patient will be censored at the last tumor scan prior to the start of ACT. However, PFS events occurring more than 90 days after discontinuation of study drug, but prior to start of subsequent ACT, will not be included in

the analysis and such patients will be censored at their last scan prior to the end of the 90-day window.

PFS will be summarized in a Kaplan-Meier plot; however in the case of a low censoring rate, it may be summarized with descriptive statistics.

9.2.3 Overall Survival

OS is defined as the date from first dose of rucaparib to the date of death due to any cause, plus 1 day. For patients who have not died as of the time of data cut-off for analysis, data will be censored at the date the patient was last known to be alive.

OS will be summarized in a Kaplan-Meier plot; however in the case of a low censoring rate, it may be summarized with descriptive statistics.

9.3 Exploratory Efficacy Analyses

The activity of rucaparib will be evaluated in molecularly-defined HRD subgroups as described below. For all exploratory efficacy analyses, ORR, PFS, and OS (as defined in Section 9) may be used as endpoints. However, not all endpoints and analyses may be performed if the subgroups become small and limit the interpretation of the results. In addition, not all samples may be tested if it becomes clear that the analysis will have limited scientific value.

9.3.1 Correlation of Biomarkers with Response to Rucaparib

In addition to the pre-specified cutpoint of 10% genome-wide LOH defining the HRD-positive population, the clinical activity of rucaparib will be evaluated in HRD-positive and HRD-negative groups defined by cutpoints across the range of all observed percent LOH values. Tumors will be analyzed by NGS to quantify the percentage of the genome with LOH and classified as HRD-positive and HRD-negative groups according to the LOH cutpoint. Outcome benefit in both groups will be evaluated.

Tumors with a deleterious alteration in a gene involved in homologous recombination repair (HRR), such as *BRCA1* and *BRCA2,* have been associated with rucaparib activity. The clinical activity of rucaparib in patients with a deleterious monoallelic or biallelic alteration in *BRCA1, BRCA2, PALB2, RAD51C,* and/or *RAD51D*, as well as other genes involved in homologous recombination DNA repair, will be evaluated. The activity of rucaparib in other biomarker-defined subgroups (e.g., microsatellite instability, tumor mutational burden, PD-L1 expression, gene expression profiling) may also be assessed.

9.3.2 Evaluation of Circulating Cell-free Tumor DNA as a Molecular Marker of Response to Rucaparib

Deleterious alterations in genes of the HRR pathway will also be interrogated using ctDNA extracted from baseline plasma samples. The clinical activity of rucaparib will be evaluated in subgroups of patients with a HRR gene alteration, as well as in subgroups of patients with other alterations detected in the ctDNA assay.

9.3.3 Assessment of Molecular Changes Over Time in Plasma and Tumor Samples

Deleterious alterations in genes of the HRR pathway, and other biomarkers (eg, microsatellite instability, tumor mutational burden, gene expression profiling), will be interrogated in tumor tissue samples (when available) and ctDNA extracted from serially collected samples. Changes in the profiles observed in these samples will be evaluated for biomarkers of response to rucaparib.

9.3.4 Evaluation of Biomarkers Associated with Resistance to Rucaparib

Tumor tissue (when available) and plasma collected at the time of disease progression will be profiled to identify potential alterations associated with resistance to rucaparib. Samples may be tested using genomic, transcriptional, and proteomic assays. Analysis of ctDNA from serially collected blood samples may be performed to identify and characterize the emergence of mechanisms of resistance. The analysis will include, but not be limited to, reversions of HRR gene alterations that restore wild-type function and alterations in other components of DNA repair and oncogenic signaling pathways.

9.4 Exploratory Pharmacokinetic Analyses

In all patients with at least one PK sample collected, the trough plasma rucaparib PK data (C_{min}) and summary statistics of C_{min} (N, mean, StD, minimum, median, maximum, coefficient of variation [CV] %), and the mean (±StD) concentration versus time plot will be reported. The PK data and selected safety and efficacy endpoints will be included in exploratory population PK and exposure-response analyses, and the results will be reported separately.

10 STATISTICAL / ANALYTICAL ISSUES

10.1 Handling of Dropouts or Missing Data

All data will be used to their maximum possible extent, but without any imputations for missing data. All time-to-event analyses will include censoring and the rules for deriving the censored values are described in more detail under each of the time-to-event endpoint descriptions.

10.2 Pooling of Centers in Multi-Center Studies

All centers will be pooled for analysis.

10.3 Multiple Comparison / Multiplicity

No adjustments for multiple comparisons will be made.

10.4 Examination of Subgroups

Analyses of efficacy, safety, and exploratory endpoints may be analyzed by baseline demographics and clinical characteristics listed in Sections 6.1and 6.2 or for other subgroups of interest as appropriate.

10.5 Interim Analyses and Data Monitoring

A group sequential interim monitoring plan will be implemented in order to help guide the data monitoring committee (DMC) in monitoring the study. The null hypothesis is p=0.10 based on historical data in similar patient populations. The study has greater than 90% power to reject the null hypothesis at a 5% significance level if the true response rate for rucaparib is 20%.

Interim analyses will be performed when approximately 60 patients and 120 patients have complete data, defined as a documented response (PR or CR), disease progression, or at least 4 months of disease assessment data available.

If the response rate does not meet continuance criteria for the ITT population, the DMC will further evaluate the benefit: risk for study treatment, both overall and separately for patients in different HRD subgroups, and make a recommendation whether further enrollment into 1 or all groups should be discontinued.

Additional details regarding this group sequential interim monitoring plan are provided in Appendix 2.

11 SAFETY ANALYSIS

The safety analyses will be performed using the safety population.

11.1 Adverse Events

Adverse events will be classified using the MedDRA classification system. The severity of the toxicities will be graded according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) whenever possible. Treatment-emergent adverse events (TEAEs) are defined as AEs with onset date on or after the date of first dose of study medication until the date of the last study medication dose plus 28 days. Adverse events will be considered treatment-emergent if all or part of the date of onset of the AE is missing and it cannot be determined if the AE meets the definition for treatment-emergent.

The number and percentage of patients who experienced TEAEs for each system organ class and preferred term will be presented. Multiple instances of the TEAE in each system organ class and multiple occurrences of the same preferred term are counted only once per patient. The number and percentage of patients with at least one TEAE will also be summarized.

The incidence of TEAEs will also be summarized by relationship to study drug according to the following categories: “treatment-related,” or “not treatment-related”. If a patient

experiences multiple occurrences of the same AE with different relationship categories, the patient will be counted once, as a relationship category of treatment-related.

If a patient experiences multiple occurrences of the same AE with different toxicity grades, the patient will be counted once for the maximum (most severe) toxicity grade. AEs with a missing toxicity grade will be presented in the summary table with a toxicity grade of "Missing." For each toxicity grade, the number and percentage of patients with at least one TEAE of the given grade will be summarized.

The time to the first TEAE and first treatment-related TEAE that results in a dose reduction, delay, interruption, or discontinuation of study drug is defined as 1+ the number of days from the first dose of study drug to the start of the first TEAE.

Separate tables will be presented including but not limited to:

- All TEAEs;
- TEAEs by CTCAE grade;
- Grade 3 or greater TEAEs;
- Treatment-related TEAEs, overall and by categories: CTCAE grade, grade 3 or greater, resulting in dose reduction, resulting in dose interruption, resulting in dose discontinuation, outcome of death;
- Serious TEAEs, overall and by categories: age, sex, and treatment-relatedness;
- TEAEs with an outcome of death;
- TEAEs leading to discontinuation of study medication;
- TEAEs resulting in interruption of study medication; and
- TEAEs resulting in reduction, delay, or interruption of study medication.
- Time to the first TEAE that results in a reduction, delay, interruption, or discontinuation of study drug.
- Time to the first treatment-related TEAE that results in a reduction, delay, interruption, or discontinuation of study drug.

Non-TEAEs (pre-treatment and post-treatment) will be presented in the by-patient data listings for the safety population.

11.2 Clinical Laboratory Evaluations

Clinical laboratory evaluations include the continuous variables for hematology, serum chemistry, and urinalysis. The laboratory values will generally be presented in International System (SI) units. The on-treatment period will be defined as the time from first dose to 28 days after the last dose of study drug. Laboratory values collected during the on-treatment period will be included in the summary tables. The laboratory values collected after the on-treatment period will only be presented in the data listings.

The summary of laboratory data will include shift tables based on CTCAE for shifts in grade from baseline to maximum, minimum, and last value during the on-treatment period.

The summary of laboratory data will include descriptive statistics (N, mean, StD, minimum, median, and maximum) of the maximum, minimum, and last value during the on-treatment period. Summaries using descriptive statistics of the change from baseline to the maximum, minimum, and last value during the on-treatment period will also be given.

Supporting laboratory data including normal ranges and abnormal laboratory flags will be provided using by-patient listings. Separate listings will be produced for clinically significant laboratory abnormalities (ie, those that meet Grade 3 or 4 criteria according to CTCAE).

11.3 Vital Signs

The on-treatment period will be defined as the time from first dose to 28 days after the last dose of study drug. Vital sign measurements collected during the on-treatment period will be included in the summary tables. The vital sign measurements collected after the on-treatment period will only be presented in the data listings.

The summary of vital sign data will include descriptive statistics (N, mean, StD, minimum, median, and maximum) of the maximum, minimum, and last value during the on-treatment period. Summaries using descriptive statistics (N, mean, StD, minimum, median, and maximum) of the change from baseline to the maximum, minimum, and last value during the on-treatment period will also be given. The data will be presented separately for each randomized treatment group and overall.

12 REFERENCES

1. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. June 1958;53(282):457-81.
2. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg. 2010;143(3):331-6.
3. Swisher EM, Lin KK, Oza AM, Scott CL, Giordano H, Sun J, et al. Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. Lancet Oncol. 2017;18(1):75-87.
4. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013;31(11):1023-31.
5. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228-47.

13 APPENDICES

Appendix 1 Response Evaluation Criteria in Solid Tumors

The RECIST guidelines (Version 1.1) are described in Eisenhauer (2009)[5] and at http://www.eortc.be/Recist/Default.htm. A short summary is given below.

Measurable Disease:

Tumor lesions: measurable lesions are defined as those that can be accurately measured in at least 1 dimension (longest diameter to be recorded) with the following:

- A minimum size of 10 mm by CT scan (CT scan thickness no greater than 5 mm).
- A minimum size of 10 mm caliper measurement by clinical exam (lesions that cannot be accurately measured with calipers should be recorded as nonmeasurable).
- A minimum size of 20 mm by chest X-ray.

All tumor measurements must be recorded in millimeters (or decimal fractions of centimeters).

Malignant lymph nodes: to be considered pathologically enlarged and measurable, a lymph node must be ≥ 15 mm in short axis when assessed by CT scan (CT scan slice thickness recommended to be not greater than 5 mm). At baseline and in follow-up, only the short axis will be measured and followed.

Nonmeasurable Disease:

All other lesions (or sites of disease), including small lesions (longest diameter < 10 mm or pathological lymph nodes with ≥ 10 to < 15 mm short axis), as well as truly nonmeasurable lesions, are considered nonmeasurable disease. Lesions considered truly nonmeasurable include leptomeningeal disease, ascites, pleural/pericardial effusions, inflammatory breast disease, lymphangitic involvement of skin and lung, and abdominal masses/abdominal organomegaly identified by physical exam that is not measurable by reproducible imaging techniques.

Bone Lesions

Bone lesions, cystic lesion, and lesions previously treated with local therapy require particular comment. Bone scan, PET scan, or plain films are not considered adequate imaging techniques to measure bone lesions. However, these techniques can be used to confirm the presence or disappearance of bone lesions.

Lytic bone lesions or mixed lytic–blastic lesions with identifiable soft tissue components that can be evaluated by cross-sectional imaging techniques such as CT or MRI can be considered as measurable lesions if the soft tissue component meets the definition of measurability described above.

Blastic bone lesions are nonmeasurable.

Cystic Lesions

Lesions that meet the criteria for radiographically defined simple cysts should not be considered as malignant lesions (neither measurable nor nonmeasurable) because they are, by definition, simple cysts.

Cystic lesions thought to represent cystic metastases can be considered as measurable lesions if they meet the definition of measurability described above. However, if noncystic lesions are present in the same patient, these are preferred as target lesions.

Lesions with Prior Local Treatment

Tumor lesions situated in a previous irradiated area or in an area subjected to other locoregional therapy are usually not considered measurable unless there has been demonstrated progression in the lesion.

Target Lesions

All measurable lesions up to a maximum of 2 lesions per organ and 5 lesions in total, representative of all involved organs, should be identified as target lesions and recorded and measured at baseline. Target lesions should be selected on the basis of their size (lesions with the longest diameter) and their suitability for accurate repeated measurements (either by imaging techniques or clinically). A sum of the longest diameter (LD) for all target lesions will be calculated and reported as the baseline sum LD. The baseline sum LD will be used as reference by which to characterize the objective tumor response.

Nontarget Lesions

RECIST criteria require unequivocal quantification of the changes in tumor size for adequate interpretation of the sum of target lesions. Consequently, when the boundaries of the primary are difficult to delineate, this tumor should not be considered a target lesion.

Guidelines for Evaluation of Measurable Disease

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Imaging-based evaluation is preferred to evaluation by clinical examination when both methods have been used to assess the antitumor effect of a treatment.

Evaluation of Target Lesions	
Complete Response	Disappearance of all target lesions. Any pathological lymph nodes (whether target or nontarget) must have reduction in short axis to < 10 mm.
Partial Response	At least a 30% decrease in the sum of the LD of target lesions, taking as reference the baseline sum LD.
Stable Disease	Neither sufficient shrinkage to qualify for partial response nor sufficient increase to qualify for PD, taking as reference the smallest sum LD since the treatment started.
Progressive Disease	At least a 20% increase in the sum of the LD of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. The appearance of one or more new lesions is also considered progression.

Abbreviations: LD = longest diameter; PD = progressive disease.

Evaluation of Nontarget Lesions	
Complete Response	Disappearance of all nontarget lesions and normalization of tumor marker level.
Stable Disease/Incomplete Response	Persistence of 1 or more nontarget lesion(s) or/and maintenance of tumor marker level above the normal limits.
Progressive Disease	Appearance of 1 or more new lesions and/or unequivocal progression of existing nontarget lesions.

If tumor markers are initially above the institutional ULN, they must normalize for a patient to be considered a complete responder.

Evaluation of Best Overall Response

The best overall response is the best response recorded from the start of the treatment until disease progression/recurrence (taking as reference for PD the smallest measurements recorded since the treatment started). The patient's best response assignment will depend on the achievement of both measurement and confirmation criteria.

Evaluation of Best Overall Response			
Target Lesions	**Nontarget Lesions**	**New Lesions**	**Overall Response**
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR

Evaluation of Best Overall Response			
Target Lesions	**Nontarget Lesions**	**New Lesions**	**Overall Response**
PR	Non-PD or not evaluated	No	PR
SD	Non-PD or not evaluated	No	SD
Not Evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR	=	complete response
NE	=	not evaluable.
PD	=	progressive disease
PR	=	partial response
SD	=	stable disease

Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be classified as having symptomatic deterioration. Every effort should be made to document the objective progression, even after discontinuation of treatment.

In some circumstances, it may be difficult to distinguish residual disease from normal tissue. When the evaluation of CR depends on this determination, it is recommended that the residual lesion be investigated (fine needle aspiration/biopsy) prior to confirming the complete response status.

Confirmatory Measurement/Duration of Response

Confirmation

If a complete response (CR) or partial response (PR) is noted, confirmatory scans should be performed at least 4 weeks after response was first documented.

Duration of Overall Response

The duration of overall response is measured from the time measurement criteria are met for CR or PR (whichever is first recorded) until the first date that recurrent or PD is objectively documented (taking as reference for PD the smallest measurements recorded since the treatment started).

The duration of overall CR is measured from the time measurement criteria are first met for CR until the first date that recurrent disease is objectively documented.

Duration of Stable Disease

Stable disease (SD) is measured from the start of the treatment until the criteria for progression are met, taking as reference the smallest measurements recorded since the treatment started.

Appendix 2 Group Sequential Interim Monitoring Plan

Introduction

This plan describes an adaptive study design for the evaluation of rucaparib in patients with locally advanced or metastatic urothelial carcinoma. The adaptive design utilizes interim analyses that allow for early stopping for futility after DMC review of futility analysis of the ITT population, the efficacy for HRD subgroups and all safety analyses.

The trial will enroll 60-200 subjects in a single experimental arm.

Objective response rate by RECIST v1.1 as assessed by the investigator will be summarized for the ITT and HRD-positive populations as well as for the exploratory HRD groups.

Primary Analysis

Let p be the true response rate. The primary analysis compares p to a performance goal of 0.10, and is thus a hypothesis test of

$$H_0 : p \leq 0.10 \qquad H_1 : p > 0.10$$

Starting with a noninformative Beta(0.1,0.9) prior on p, the posterior distribution is

$$p|data \sim Beta(0.1+Y,\ 0.9+N-Y)$$

where Y is the observed number of responses and N is the number of complete subjects in the trial. The trial is a success if

$$Pr(p > 0.10 \mid data) > 0.95$$

This threshold maintains type I error at 5%. We consider an alternative threshold (0.935) that increases type I error slightly but obtains greater power for comparison.

Design

The trial sequentially enrolls patients and interim analyses are conducted. At the time of each interim analysis, there will be N patients with a 4-month endpoint and M patients that are incomplete. Let Y be the number of responders among the complete patients.

The model computes p_{max}, the predictive probability of trial success if the trial is followed to the maximum (targeted) sample size and all patients are followed to completion. Let W_{max} be the number of eventual responders among the unknown patients (incomplete and currently unenrolled). The distribution of W_{max} is a BetaBinomial(MAXN-N,0.1+Y,0.9+N-Y). The predictive probability p_{max} is the probability W_{max} results in trial success when N=200 subjects are complete.

The model indicates trial futility at any interim analysis if p_{max}<threshold, indicating that there is limited probability that the trial will reach success even at the maximum sample size.

For this study design with N=60 minimum, N=200 maximum, a 5% futility threshold, and interims at N=60 and N=120, the required number of successes to continue is summarized in the table below. For example, in the first row we require 5/60, indicating we must have at least 5 successes out of the first 60 patients (so 5 would continue).

Interim	Minimum to Continue (5% predictive probability)
1 (N=60)	5/60 (8.3%)
2 (N=120)	13/120 (10.8%)

Example Trials

Example trial 1

The trial enrolls until the first interim analysis at N=60 ITT patients complete. At this interim, we have 13/60 (21.7%) responses with 40 patients awaiting follow-up. This result is clearly encouraging compared to a 10% baseline rate. With 140 patients still requiring data, the predictive probability of success on the ITT population is 97.5%, well above the threshold for futility. Response among HRD subgroups is evaluated and the trial continues.

The second interim analysis occurs with N=120 ITT patients complete, obtaining 22/120 (18.3%) with 27 incomplete. Response among HRD subgroups is evaluated. With 80 patients remaining needing follow-up, the predictive probability of success is 99.0%, clearly high and the trial continues.

At N=200, we obtain 37/163 responses (22.7%). This results in a posterior probability greater than 99.99% and a successful trial.

Example Trial 2

The trial enrolls until the first interim analysis at N=60 ITT patients complete. At this interim, we have 3/60 (5%) responses with 40 patients awaiting follow-up. The predictive probability of success at this point is only 0.4%. This is below the threshold for futility for the ITT population. The DMC would examine response among HRD subgroups as well as safety data and make a recommendation to stop the trial or continue. If the trial was stopped at this point, it is assumed that all enrolled patients will continue to be followed.

In that follow-up, the trial ended with 12/100 responses (12%). While greater than 10%, this is insufficient for a successful trial (Pr(trmt beats ctrl)=56.7%) and the trial would have been very unlikely to meet its primary objective even had N=200 subjects been enrolled.

Example Trial 3

The trial enrolls until the first interim analysis at N=60 ITT patients complete. At this interim, we have 5/60 (8.3%) responses with 40 patients awaiting follow-up. With 140 patients still requiring data, the predictive probability of success on ITT is 5.1%, barely above the threshold for futility. Response among HRD subgroups is evaluated and the trial continues.

The second interim analysis occurs with N=120 complete, obtaining 14/120 (11.7%) with 31 incomplete. With 80 patients remaining needing follow-up, the predictive probability of success is 13.1%, also high enough to continue. Response among HRD subgroups is evaluated and the DMC recommends that the trial continue.

At N=200, we obtain 20/200 responses (10.0%). This results in a posterior probability of 47.5% and a nonsuccessful trial.

Simulations and Operating Characteristics

The operating characteristics of this trial were determined through trial simulation and compared to a potential Simon 2-stage design. We computed the operating characteristics under the assumptions that p=0.10 (null hypothesis) and p=0.11, 0.12, …, 0.25.

A total of 10,000 trials were simulated per scenario. In each, subjects were accrued according to a Poisson process with an average of 1.5 subjects per week. In each simulated trial, interims analyses were conducted according to prespecified rules and the results were recorded.

The detailed results are shown in the following tables for Expected Sample size and for Power, respectively.

Expected Sample Size	p=0.10 (null)	p=0.15	p=0.20	p=0.25
Fixed	200.00	200.00	200.00	200.00
Simon	141.52	188.83	204.28	205.92
Adjusted	**150.85**	**192.38**	**199.60**	**199.98**
Unadjusted	151.60	192.67	199.40	199.99

Pr(trial success)	p=0.10 (null)	p=0.15	p=0.20	p=0.25
Fixed	0.043	0.683	0.989	0.999
Simon	0.049	0.673	0.977	0.999
Adjusted	**0.062**	**0.736**	**0.991**	**1.000**
Unadjusted	0.039	0.675	0.985	0.999

The next table shows the details of if and where each of the adjusted adaptive trial simulations stopped for futility. For each simulated trial, the trial could stop at N=60 complete, N=120, or run to the maximum sample size. Note that trials that stop at N=60 or N=120 complete actually we enroll more than 60 or 120 subjects due to the "overrun", patients enrolled while waiting for the required patient to complete. In addition, we summarize the probability of a "futility to success flip flop" (FSFF). In actual implementation, when a trial is stopped for futility, you will not get to see what would have happened had you enrolled until N=200. However, in simulation we can simulate those remaining subjects and determine how many futility stops occurred for trials that would ultimately have been successful.

Pr(trial success)	p=0.10 (null)	p=0.15	p=0.20	p=0.25
Futility N=60	0.270	0.042	0.004	0.000
Futility N=120	0.331	0.054	0.000	0.000
Pr(FSFF)	0.004	0.022	0.004	0.000