JAVELIN OVARIAN PARP 100

A RANDOMIZED, OPEN-LABEL, MULTICENTER, PHASE 3 STUDY TO EVALUATE THE EFFICACY AND SAFETY OF AVELUMAB IN COMBINATION WITH CHEMOTHERAPY FOLLOWED BY MAINTENANCE THERAPY OF AVELUMAB IN COMBINATION WITH THE POLY (ADENOSINE DIPHOSPHATE [ADP]-RIBOSE) POLYMERASE (PARP) INHIBITOR TALAZOPARIB IN PATIENTS WITH PREVIOUSLY UNTREATED ADVANCED OVARIAN CANCER

STATISTICAL ANALYSIS PLAN – B9991030

Compounds:	MSB0010718C MDV3800, BMN 673
Compound Name:	Avelumab Talazoparib
Version:	1.0
Date:	29-May-2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B9991030 is based on the protocol dated 30MAR2018.

Table 1. Summary of Major Changes in SAP Amendments

Version	Version Date	Summary of Changes
1	29-May-2018	Not applicable (N/A)

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B9991030. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

A separate SAP will cover the interim analyses for periodic safety review by the External Data Monitoring Committee (E-DMC).

Statistical analyses will be performed using cleaned eCRF data as well as non-CRF data (ie, pharmacokinetic (PK) concentration, anti-drug antibodies (ADA; neutralizing antibody [nAb]), supplemental biomarker data, and tumor assessment results by the Blinded Independent Central Review [BICR]). The primary analysis will include all data up to a cut-off date which is determined by the number of events for progression-free survival (PFS) by BICR in patients with advanced ovarian cancer with defects in DNA damage repair（DDR+) and minimum follow-up of 12 months. The cut-off date is determined once a data extract (before database lock) is available which indicates that the required number of events for PFS by BICR in the DDR+ population and minimum follow-up of 12 months is expected to occur by the cut-off date.

Due to cleaning activities, the final number of events might deviate from the planned number. The data cut-off date will not be adjusted retrospectively in this case.

2.1. Study Objectives

Primary Objective

To demonstrate that avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance is superior to platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance in prolonging PFS in patients with advanced ovarian cancer with defects in DNA damage repair (DDR+).

Secondary Objectives

- To demonstrate that avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance is superior to platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance in prolonging overall survival (OS) in patients with advanced ovarian cancer with defects in DNA damage repair (DDR+).

- To demonstrate that avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance is superior to platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance in prolonging PFS in patients with advanced ovarian cancer unselected for DDR status.
- To demonstrate that avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance is superior to platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance in prolonging OS in patients with advanced ovarian cancer unselected for DDR status.
- To evaluate the effect on PFS and OS of platinum-based chemotherapy followed by talazoparib maintenance versus platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance in patients with advanced ovarian cancer with DDR+ and unselected for DDR status.
- To evaluate the anti-tumor activity in each treatment arm.
- To evaluate the overall safety profile in each treatment arm.
- To evaluate the PK of avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance in Arm A as well as well as PK of talazoparib in combination with avelumab (Arm A) and as a single agent (Arm B).
- To evaluate the immunogenicity of avelumab in Arm A (avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance and avelumab in combination with platinum-based chemotherapy).
- To evaluate the effect of avelumab in combination with platinum-based chemotherapy followed by avelumab plus talazoparib maintenance and platinum-based chemotherapy followed by talazoparib maintenance versus platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance on patient reported outcome (PRO) including the assessments of disease-related symptoms, treatment side effects and health-related quality of life (HRQoL).
- To assess the correlation of anti-tumor activity with PD-L1 expression, tumor mutational burden (TMB) and with potential biomarkers of PARP inhibitor sensitivity in baseline tumor tissue.
- To assess the correlation of anti-tumor activity with TMB and potential biomarkers of PARP inhibitor sensitivity in baseline circulating tumor (ct) DNA.

CCI [redacted]

- [redacted]

- [redacted]

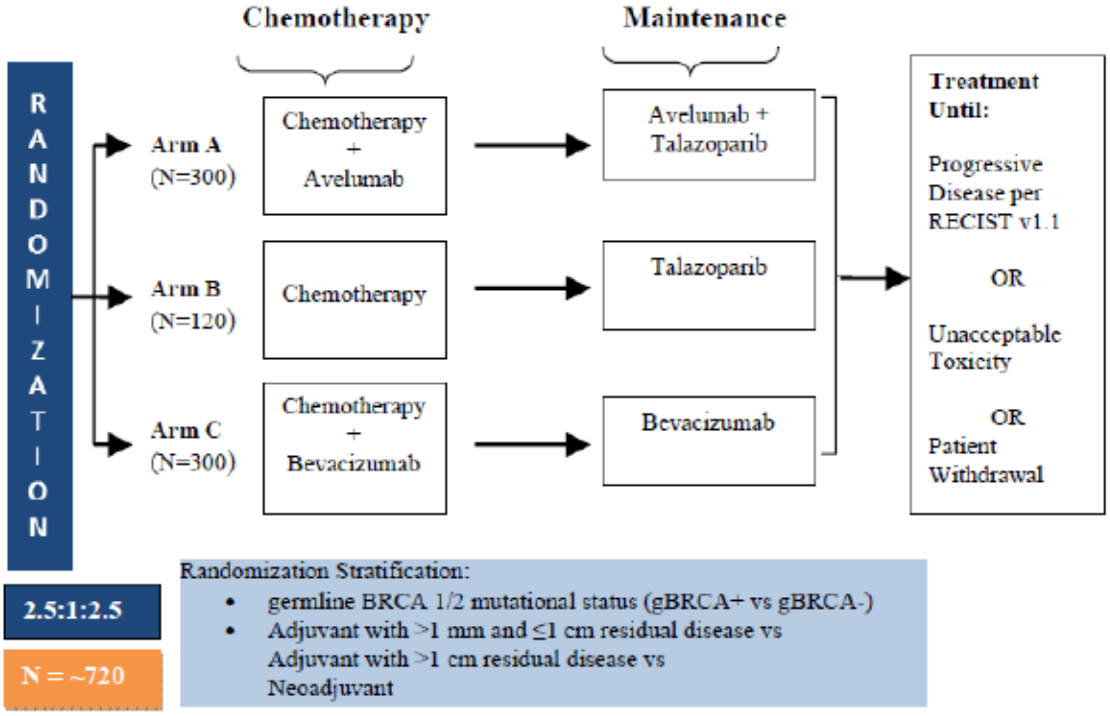
[redacted]

2.2. Study Design

This is a Phase 3, randomized, open-label, multicenter study to evaluate the efficacy and safety of avelumab in combination with chemotherapy followed by maintenance therapy of avelumab in combination with the PARP inhibitor talazoparib in patients with previously untreated advanced ovarian cancer.

The study design is illustrated in the following figure.

Figure 1. Study Design Schema



gBRCA = germline BReast CAncer gene; N = number of patients

Crossover between treatment arms will not be permitted. Randomization will be stratified by germline BRCA 1/2 mutational status (gBRCA+ vs gBRCA-) and resection status (adjuvant with >1 mm and ≤1 cm residual disease vs adjuvant with >1 cm residual disease vs neoadjuvant).

Intravenous carboplatin and paclitaxel will be used as the chemotherapy backbone, consisting of every 3 week (Q3W) carboplatin and paclitaxel.

Patients may be enrolled either following primary debulking surgery (PDS), or prior to initiation of neoadjuvant chemotherapy. The latter group will undergo interval debulking

surgery (IDS) after 3 study cycles of chemotherapy (plus avelumab in Arm A, or plus bevacizumab in Arm C) to be followed, upon recovery from surgery, by the remainder of chemotherapy (plus avelumab in Arm A, or plus bevacizumab in Arm C).

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

- PFS as determined based on BICR assessment per RECIST v1.1 in patients with newly diagnosed advanced ovarian cancer with defects in DNA damage repair (DDR+).

  PFS is defined as the time from the date of randomization to the date of the first documentation of progressive disease (PD) or death due to any cause, whichever occurs first.

3.2. Secondary Endpoints

3.2.1. Safety endpoints

- AEs (as graded by National Cancer Institute [NCI] Common Terminology Criteria for Adverse Events [CTCAE] v4.03); laboratory abnormalities (as graded by NCI CTCAE v4.03); vital signs (blood pressure, pulse rate); electrocardiograms (ECGs).

  AEs will be graded by the investigator according to CTCAE v4.03 and coded using the Medical Dictionary for Regulatory Activities (MedDRA).

3.2.2. Efficacy endpoints

- OS in patients with tumors that are DDR+ and in patients unselected for DDR status.

  OS is defined as the time from the date of randomization to the date of death due to any cause.

- PFS based on BICR assessment per RECIST v1.1 in patients unselected for DDR status.

- PFS based on investigator assessment per RECIST v1.1 in patients with tumors that are DDR+ and in patients unselected for DDR status.

- PFS2 based on investigator assessment in patients with tumors that are DDR+ and in patients unselected for DDR status.

  PFS2 is defined as time from the date of randomization to the start of second subsequent treatment after first documentation of PD, or death from any cause, whichever occurs first.

- PFS based on investigator assessment per Gynecological Cancer Intergroup (GCIG) criteria in patients with tumors that are DDR+ and in patients unselected for DDR status.

  PFS by GCIG criteria will be assessed in this study incorporating both RECIST v1.1 and CA-125 (Rustin, G et.al, 2011).

3.2.3. Patient reported outcomes

- Disease related symptoms and treatment side effects as measured by the NCCN-FACT FOSI-18 and health-related quality of life (HRQOL) as measured by NCCN-FACT FOSI-18 and the EuroQol Group 5-Dimension 5-Level (EQ-5D-5L).

  NFOSI-18 (Jensen et al. 2011, Jensen et al. 2015) is an ovarian cancer-specific symptom index comprised of symptoms rated as highest priority by both oncology clinical experts and women with advanced ovarian cancer. The NFOSI-18 was developed to be part of the Functional Assessment of Chronic Illness Therapy (FACIT) system and was specifically created with the input from the Food and Drug Administration (FDA), including the recommendation that the assessment of specific symptoms is an appropriate starting point for improved measurement of QoL domains. It is specifically designed to be a stand-alone instrument to measure disease-related symptoms, treatment side effects and function/well-being in patients with ovarian cancer.

  The disease-related symptoms subscale is composed of the disease related symptoms-physical (DRS-P) domain which included 9 physical symptoms/concerns (energy, pain, ill, stomach cramps, fatigue, constipation, stomach swelling, bowel control, and sleep) and the disease related symptoms-emotional (DRS-E) which included 1 emotional symptom/concern (worry condition will get worse). The treatment side effect (TSE) subscale included 5 items (nausea, hair loss, bothered by side effects, vomiting, and skin problems). The functional well-being (FWB) subscale included 3 items (able to get around, enjoy life, and content with QoL). As with all FACIT questionnaires, a high score is good. Therefore, a score of "0" is a severely symptomatic patient and the highest possible score is an asymptomatic patient.

  The EuroQol EQ-5D-5L (Herdman at al. 2011; Rabin and de Charro 2001) is a 6 item patient-completed questionnaire designed to assess health status in terms of a single index value or utility score. There are 2 components to the EQ-5D-5L, a Health State Profile which has individuals rate their level of problems (none, slight, moderate, severe, extreme/unable) in 5 areas (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and a Visual Analogue Scale (VAS) in which patients rate their overall health status from 0 (worst imaginable) to 100 (best imaginable). Published weights are available that allow for imputation of the index score. Overall index scores range from 0 to 1, with low scores representing a higher level of dysfunction.

3.2.4. Pharmacokinetic endpoints

- PK parameters, including C_{trough} and C_{max} for avelumab and C_{trough} for talazoparib.

  PK parameters [C_{max} and C_{trough} for avelumab (Arm A) and C_{trough} for talazoparib (Arms A and B)] will be reported after single dose and at steady state. For talazoparib, dose normalized C_{trough} [C_{trough}(dn)] will also be determined, as appropriate.

Table 2. PK Parameters to be Determined for Avelumab and Talazoparib

Parameter	Definition	Method of Determination
C_{max}	Maximum observed plasma concentration for avelumab	Observed directly from data
C_{trough}	Predose concentration during multiple dosing	Observed directly from data
$C_{trough}(dn)$	Dose normalized C_{trough} for talazoparib	C_{trough} / Dose

3.2.5. Immunogenicity endpoints

- Anti-drug antibodies (ADA) and neutralizing antibody (nAb) against avelumab.

3.2.6. Biomarker endpoints

- PD-L1 expression, TMB, genomic scarring and the presence of defects in select genes, considered critical to effective DDR, in baseline tumor tissue.
- Assessment of defects in a panel of key oncogenes, including several considered critical to effective DDR and TMB in ctDNA at baseline.

Table 3. Biomarker Definition and Determination

Parameter	Definition	Method of Determination
PD-L1 expression	The number of PD-L1 positive cells and/or qualitative assessment of PD-L1 staining on tumor and inflammatory cells in regions of interest that are defined by tumor cell morphology and the presence or absence of inflammatory cells	Pathologist, assisted by image analysis
Genomic scarring and the Presence of defects in select genes, considered critical to effective DDR, in baseline tumor tissue	Quantitation of genomic scarring in the form of loss of heterozygosity; The number of somatic and germline mutations present in a panel of genes associated with DDR in baseline tumor derived nucleic acid, in germline nucleic acid and in circulating tumor DNA.	Next generation sequencing followed by computational analysis
TMB	Determination/estimation of the frequency of mutations (total and non-synonymous) present in baseline tumor derived nucleic acid samples and in baseline circulating tumor DNA.	Whole exome or genome sequencing and/or RNAseq

CCI [redacted]

[redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

[redacted]

3.4. Baseline Variables

3.4.1. Study drug, study treatment, and baseline definitions

In this study, '**study drug**' refers to avelumab, paclitaxel, carboplatin, bevacizumab, or talazoparib and '**study treatment**' (or '**treatment arm**') refers to one of the following:

- Arm A = platinum-based chemotherapy + avelumab followed by avelumab + talazoparib maintenance;
- Arm B = platinum-based chemotherapy followed by talazoparib maintenance;
- Arm C = platinum-based chemotherapy + bevacizumab followed by bevacizumab maintenance.

Start and end dates of study treatment:

The date/time of first dose of study treatment in a combination arm is the earliest date/time of the first non-zero dose date/time for the study drugs in the combination.

The date/time of last dose of study treatment in a combination arm is the latest date/time of the last non-zero dose date/time for the study drugs in the combination.

Definition of baseline:

Definition of baseline for efficacy and PRO analyses

The last measurement prior to randomization will serve as the baseline measurement for efficacy and PRO analyses. If such a value is missing (since per protocol the first PRO assessment is planned to occur prior to dosing on Cycle 1 Day 1), the last measurement prior to the first dose of study treatment will be used as the baseline measurement except for analyses of tumor assessments data where the baseline assessment would be considered as missing.

Definition of baseline for immunogenicity analyses

The last available assessment prior to the start of treatment with avelumab is defined as 'baseline' result or 'baseline' assessment. If an assessment is planned to be performed prior to the first dose of avelumab in the protocol and the assessment is performed on the same day as the first dose of avelumab, it will be assumed that it was performed prior to avelumab administration, if assessment time point is not collected or is missing.

Definition of baseline for safety analyses

The last available assessment prior to the start of study treatment is defined as 'baseline' value or 'baseline' assessment for safety analyses. If an assessment is planned to be performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration, if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 for baseline calculation. Unscheduled assessments will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

Patients who start treatment and discontinue from the study on the same day may have two different sets of data collected on study day 1 (one during study and one in the End of Treatment (EOT) visit. Data reported at the EOT visit are not eligible for baseline selection.

If a scheduled pre-dose measurement actually occurred post-dose, then the corresponding measurement will be treated and analyzed similar to an unscheduled post-dose measurement.

Definition of baseline for biomarker analyses

The last available assessment prior to the start of study treatment is defined as 'baseline' value or 'baseline' assessment for biomarker analyses. For biomarkers that are planned to be measured on Cycle 1 Day 1 (eg, soluble proteins), if the assessment time point is not collected or is missing, it will be assumed that the measurement was performed prior to first dose of study treatment.

3.4.2. Baseline characteristics

Randomization is stratified by the following, as recorded in the Interactive Response Technology (IRT):

- Germline BRCA 1/2 mutational status (gBRCA+ vs gBRCA-)
- Resection status (adjuvant with >1 mm and ≤1 cm residual disease vs adjuvant with >1 cm residual disease vs neoadjuvant).

The primary analyses of PFS in the DDR+ population, as well as secondary analyses of OS in the DDR+ population, and PFS and OS in patients unselected for DDR status will be stratified by these randomization stratification factors.

Patients will be defined as having defective DDR (DDR+) or having intact DDR (DDR-) using an analytically validated, investigational use only (IUO)-labelled next generation sequencing-based assay developed by Foundation Medicine. The test and its associated technology will be executed by Foundation Medicine at their design controlled facility as a single centralized site. The assay will assess baseline formalin fixed paraffin embedded tumor tissue from each patient for defects in the BRCA1 and BRCA2 genes as well as for the

degree of loss of heterozygosity (LOH) present. Patients' tumors will be defined as DDR+ if they have a defect in either BRCA1 or BRCA2 or if they have an LOH score above a predefined and analytically validated cut-off. All other patients' tumors will be defined as DDR-.

Other baseline characteristics (including demographics, physical measurements, disease history, and prior anti-cancer surgery) are described in Section 6.5.1. These baseline characteristics are not planned to be included as stratification variables or covariates in statistical models unless otherwise specified in Section 5.3.3.5.

3.5. Safety Endpoints

3.5.1. Adverse events

Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period for the first time, or if the worsening of an event is during the on-treatment period.

On-treatment period is defined as the time from the first dose of study treatment through minimum (30 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day). The start day of new anti-cancer drug therapy after the first dose of study treatment is derived as outlined in Section 5.2.5.

Adverse Events of Special Interest (AESIs)

AESIs are immune-related adverse events (irAE) and infusion-related reactions (IRRs). The criteria for classification of an AE as an irAE or IRR are described in Appendix 1 and Appendix 2, respectively.

3-Tier Adverse Events

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analyses are generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates these exploratory analyses.

Adverse events and clusters of adverse events, of any causality and treatment-related, will also be summarized following a 3-tier approach. Under this approach, AEs are classified into 1 of 3 tiers.

Tier-1 events: These are pre-specified events or clusters of events of clinical importance and will be described in the Safety Review Plan.

Tier-2 events: These are events that are not Tier-1 but are "common". A MedDRA PT is defined as a Tier-2 event if it is reported by

a) at least 10% of patients with any grade in any treatment arm, or

b) at least 5% of patients with Grade 3, 4 or 5 in any treatment arm.

Tier-3 events: All other AEs that are classified neither as Tier-1 nor Tier-2.

4. ANALYSIS SETS

Data for all patients will be assessed to determine if patients meet the criteria for inclusion in each analysis population prior to releasing the database and classifications will be documented per Pfizer's standard operating procedures.

Only patients who signed informed consent will be included in the analysis sets below.

Table 4 summarizes the use of the analysis sets for efficacy, safety, baseline characteristics, and exposure.

Table 4. Statistical Analyses by Analysis Set

Endpoints	Full Analysis Set	Per Protocol Analysis Set	Safety Analysis Set
Baseline Characteristics	✓		✓
Prior and Concomitant Therapies	✓		✓
Exposure			✓
Efficacy: Primary	✓	✓	
Efficacy: Secondary	✓	(key secondary endpoints only*)	
Efficacy: Exploratory	✓		
Safety			✓

* Key secondary endpoints are OS in the DDR+ population, and PFS based on BICR assessment per RECIST v1.1 and OS in patients unselected for DDR status.

4.1. Full Analysis Set

The full analysis set (FAS) will include all randomized patients. Patients will be classified according to the study treatment assigned at randomization.

4.2. Safety Analysis Set

The safety analysis set will include all patients who receive at least one dose of study drug. Patients will be classified according to the study treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case patients will be classified according to the first study treatment received.

4.3. Other Analysis Set

4.3.1. Per-protocol analysis set

Per protocol (PP) analysis set is a subset of the FAS and will include patients who do not meet any of the following criteria that could impact the key objectives of the study.

Patients who meet any of the following criteria will be excluded **from the PP analysis set for PFS by BICR**.

- Patient did not receive at least one dose of chemotherapy in the chemotherapy period
- Patient randomized to Arm A who did not receive at least one dose of avelumab
- Patient randomized to Arm C who did not receive at least one dose of bevacizumab
- Patient received a study drug to which he/she was not assigned at randomization
- Patient without a tumor assessment on or after Week 9 (Day 63) from randomization (unless PD by BICR or death is observed before that time in which case the patient will not be excluded from the PP analysis set)
- ECOG status ≥ 2 on or prior to randomization date
- Patients did not meet inclusion criteria 1, 2, or 3

    1. Histologically confirmed Stage III-IV epithelial ovarian, fallopian tube, or primary peritoneal cancer (according to American Joint Committee on Cancer (AJCC)/UICC TNM and International Federation of Gynecology and Obstetrics (FIGO) Staging System 2014 edition), including carcinosarcoma with high-grade serous component.

    2. Patients must be candidates for bevacizumab in combination with platinum based chemotherapy and previously untreated.

    3. Must have completed a primary surgical debulking procedure, or be candidates for neoadjuvant chemotherapy with planned interval debulking surgery.

        a. Patients who completed primary debulking must have had incompletely resected disease that is macroscopically/grossly visible and at least with lesions >1 mm and be randomized at a maximum of 8 weeks after surgery.

        b. For patients who are candidates for neoadjuvant chemotherapy, the diagnoses must have been confirmed by:

            - Core tissue (not fine-needle aspiration) biopsy is required for diagnosis. The tissue must be consistent with inclusion criteria #1 above.

            - Stage IIIC–IV documented via imaging or surgery (without attempt at cytoreduction).

            - Serum CA-125/CEA ratio >25. If the serum CA-125/CEA ratio is <25, then workup should be negative for the presence of a primary gastrointestinal or breast malignancy (<6 weeks before start of neoadjuvant treatment).

            - Randomization must occur within 8 weeks after diagnosis.

Patients who meet any of the following criteria will be excluded from the **PP analysis set for OS**.

- Patients did not receive at least one dose of chemotherapy in the chemotherapy period
- Patient randomized to Arm A who did not receive at least one dose of avelumab
- Patients randomized to Arm C who did not receive at least one dose of bevacizumab
- Patient received a study drug to which he/she was not assigned at randomization
- ECOG status ≥ 2 on or prior to randomization date
- Patients did not meet inclusion criteria 1, 2, or 3

4.3.2. PK analysis sets

The PK concentration analysis set is a subset of the safety analysis set and will include patients who have at least one post-dose concentration measurement above the lower limit of quantitation (LLQ) for avelumab (Arm A) or talazoparib (Arms A and B). On this study the PK parameters are observational, and therefore the PK parameter analysis set is identical to the PK concentration analysis set.

4.3.3. Biomarker analysis sets

The biomarker analysis set for biomarkers that are measured only at screening is a subset of the safety analysis set and will include patients who have at least one screening biomarker assessment.

The biomarker analysis set for biomarkers that are measured sequentially is a subset of the safety analysis set and will include patients who have at least one baseline and one on-treatment biomarker assessment for the same marker. The biomarker analysis set is defined separately for each biomarker of interest.

Analysis sets will be defined separately for blood-based and tumor tissue-based biomarkers.

4.3.4. Immunogenicity analysis set

The immunogenicity analysis set is a subset of the safety analysis set and will include patients who have at least one ADA/nAb sample collected for avelumab in the avelumab containing arm (Arm A).

CCI [redacted]

[redacted].

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

All references to PFS in this section pertain to PFS based on BICR assessment even when not specifically stated.

5.1.1. Hypotheses and sample size determination

The following statistical hypothesis will be tested to address the primary objective:

H_{01}: $HR_{PFS+} \geq 1$ vs H_{11}: $HR_{PFS+} < 1$

where HR_{PFS+} is the hazard ratio (Arm A vs Arm C) of PFS based on BICR assessment in the DDR+ population.

In addition the following statistical hypotheses will be tested to address secondary objectives:

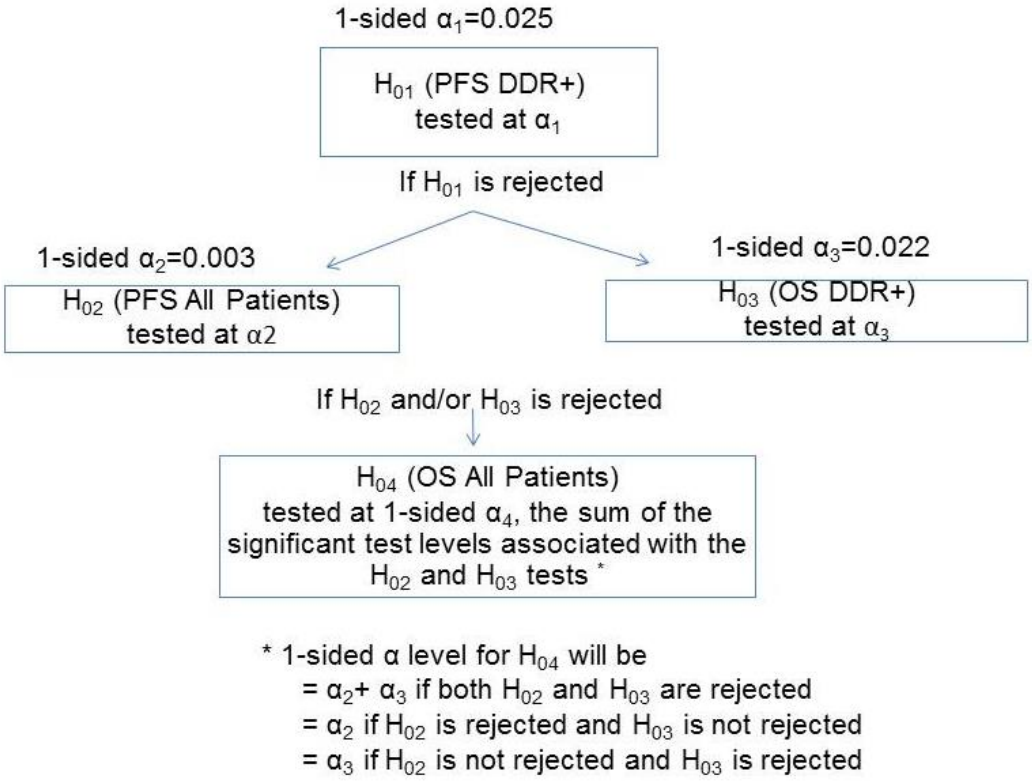
H_{02}: $HR_{PFS} \geq 1$ vs H_{12}: $HR_{PFS} < 1$

H_{03}: $HR_{OS+} \geq 1$ vs H_{13}: $HR_{OS+} < 1$

H_{04}: $HR_{OS} \geq 1$ vs H_{14}: $HR_{OS} < 1$

where HR_{PFS} and HR_{OS} are the hazard ratios (Arm A vs Arm C) of PFS based on BICR assessment and OS, respectively, in patients unselected for DDR status, and HR_{OS+} is the hazard ratio (Arm A vs Arm C) of OS in the DDR+ population. A gatekeeping procedure as illustrated in Figure 2, will be used to maintain the overall type I-error in the study at or below 1-sided 0.025. The significance levels for each test also take into account the group sequential nature of the design (see Section 5.1.2).

Figure 2. Statistical Testing Strategy for Arm A vs Arm C Comparisons



For the primary comparison of PFS between Arm A and Arm C in the DDR+ population, 200 PFS events based on BICR assessment in Arms A and C combined will provide 90% power to detect a HR of 0.63 using a 1-sided log-rank test at a significance level of 0.025, and a 2-look group sequential design with Lan-DeMets (O'Brien-Fleming) α-spending function to determine the efficacy boundary and a Gamma Family (-9) β-spending function to determine the non-binding futility boundary.

The study will randomize a total of approximately 720 patients in a 2.5:1:2.5 ratio to Arm A (N=300 including N=150 DDR+), Arm B (N=120 including N=60 DDR+), and Arm C (N=300 including N=150 DDR+). Randomization will be stratified by germline BRCA 1/2 mutational status (gBRCA+ vs gBRCA-), and resection status (adjuvant with >1mm and ≤1 cm residual disease vs adjuvant with residual disease >1cm vs neoadjuvant).

The sample size for this study is determined based on the following assumptions:

- The median PFS for patients in Arm C is 14.1 months (Burger at al. 2011), and the median PFS for patients in Arm A is 22.4 months for the DDR+ population and 20.1 months for patients unselected for DDR status; this corresponds to a hazard ratio (HR) of 0.63 in the DDR+ population and 0.70 in patients unselected for DDR status under the exponential model assumption;

- The median OS for patients in Arm C is 39.7 months (Burger at al. 2011), and the median OS for patients in Arm A is 56.7 months for the DDR+ population and 52.9 months for patients unselected for DDR status; this corresponds to a HR of 0.70 in the DDR+ population and 0.75 in patients unselected for DDR status under the exponential model assumption;

- PFS drop-out rate of approximately 15% and OS drop-out rate of approximately 5% at the time of primary analysis of PFS and OS;

- 50% of the randomized patients are with tumors that are DDR+;

- Non-uniform patient accrual accomplished over an 18-month period.

The sample size of approximately 720 patients will also allow an assessment of OS in the DDR+ population, and of PFS and OS in patients unselected for DDR status.

If H_{01} is rejected then PFS in patients unselected for DDR status and OS in the DDR+ population can be tested.

- For the PFS comparison between Arm A and Arm C in patients unselected for DDR status, 411 PFS events based on BICR assessment will provide 80.5% power to detect a HR of 0.70 using a 1-sided log-rank test at a significance level of 0.003, and a 2-look group-sequential design with Lan-DeMets (O'Brien-Fleming) α -spending function to determine the efficacy boundary.

- For the OS comparison between Arm A and Arm C in the DDR+ population, 207 deaths will provide 70.1% power to detect a HR of 0.70 or 80.1% power to detect a HR of 0.67 using a 1-sided log-rank test at a significance level of 0.022, and a 4-look group sequential design with Lan-DeMets (O'Brien-Fleming) α -spending function to determine the efficacy boundary.

If either H_{02} or H_{03} are rejected, then OS in patients unselected for DDR status can be tested at the sum of the significance levels associated with the significant H_{02} and H_{03} tests (see Figure 2). With 421 deaths, the power to detect a HR of 0.75 is 83.2% (if both H_{02} and H_{03} are rejected), 57.6% (if H_{02} is rejected and H_{03} is not rejected), or 81.9% (if H_{02} is not rejected and H_{03} is rejected) using a 1-sided log-rank test at a significance level of 0.025, 0.003, or 0.022, respectively, and a 4-look group sequential design with Lan-DeMets (O'Brien-Fleming) α-spending function to determine the efficacy boundary.

The data cut-off for the primary PFS analysis will occur after the target number of PFS events (200) based on BICR assessment in Arms A and C combined in the DDR+ population has been reached and the last patient with tumor that is DDR+ randomized in the study has been followed for at least 12 months after randomization.

The data cut-off for the primary analysis of OS will occur after the target number of deaths (207) in Arms A and C combined in the DDR+ population has been reached.

The study will be considered positive if the 1-sided stratified log-rank test for PFS based on BICR assessment for comparing Arm A to Arm C in the DDR+ population is significant at the pre-specified α level.

Arm B is not included in formal hypothesis testing. The sample size for Arm B is determined as follows.

To establish that platinum-based chemotherapy followed by talazoparib maintenance (Arm B) has proof of efficacy compared to platinum-based chemotherapy plus bevacizumab followed by bevacizumab maintenance (Arm C), a HR target of 0.792 for Arm B vs Arm C is deemed to be clinically important (corresponding to a minimum improvement of 3.7 months in median PFS over Arm C under the exponential model assumption). Proof of efficacy for Arm B compared to Arm C will have been demonstrated if both of the following criteria are met in the DDR+ population:

1. Clinical relevance: Estimated $HR_{PFS+ BvsC)} \leq 0.792$.

2. Statistical significance: Probability (true $HR_{PFS+(BvsC)} < 1 \mid$ data) $\geq 80\%$.

where $HR_{PFS+(BvsC)}$ is the HR of Arm B vs Arm C of PFS based on BICR assessment in the DDR+ population. The calculation of the statistical significance criterion assumes asymptotic normality of log ($HR_{PFS+(BvsC)}$) and a non-informative prior.

Table 5 shows the probability of demonstrating proof of efficacy for Arm B compared to Arm C under different scenarios. If the true $HR_{PFS+(BvsC)}$ is 1 (ie, no difference between the two arms) the probability of claiming proof of efficacy (false detection rate) for Arm B

compared to Arm C is 0.099. If the true $HR_{PFS+(BvsC)}$ is 0.68 and the true $HR_{PFS+(AvsC)}$ is 0.63, then 149 PFS events based on BICR assessment would be observed in the DDR+ population for Arms B and C combined at the time of the primary analysis and the probability of declaring proof of efficacy (true detection rate) for Arm B compared to Arm C is 0.80.

Table 5. Operating Characteristics for PFS in the DDR+ Population (Arm B vs Arm C)

True $HR_{PFS+(BvsC)}$	Probability of demonstrating proof of efficacy*
0.65	0.862
0.68	0.800
0.75	0.618
0.792	0.500
0.90	0.240
1.00	0.099
*Assuming 149 PFS events in the DDR+ population for Arms B and C combined at the time of the primary analysis for PFS.	

5.1.2. Decision rules

The interim and the primary analyses for each efficacy endpoint will be performed after all patients have been randomized in the study and the target number of events has occurred in Arms A and C combined as described below. A maximum of 4 distinct analyses cut-offs are planned in the study at the approximate calendar times described below:

- At the time when approximately 140 PFS events (70% of the expected 200 events) based on BICR assessment have occurred in the DDR+ population in Arms A and C combined, and all patients have been randomized in the study;
- At the time when 200 PFS events based on BICR assessment have occurred in the DDR+ population in Arms A and C combined, and the last patient randomized in the DDR+ population has been followed for at least 12 months after randomization;
- At the time when 155 deaths (74.9% of the expected 207 deaths) have occurred in the DDR+ population in Arms A and C combined;
- At the time when 207 deaths have occurred in the DDR+ population in Arms A and C combined.

A maximum of 2 analyses will be performed for PFS (in the DDR+ population and in patients unselected for DDR status):

- An interim analysis after all patients have been randomized in the study and at least 140 PFS events (70% of the expected 200 events) based on BICR assessment have occurred in the DDR+ population in Arms A and C combined;
- The final analysis for PFS after all patients in the DDR+ population have been followed for 12 months and at least 200 PFS events based on BICR assessment have occurred in the DDR+ population in Arms A and C combined;

A maximum of 4 analyses will be performed for OS (in the DDR+ population and in patients unselected for DDR status):

- An interim analysis at the time of interim analysis for PFS
- An interim analysis at the time of the final analysis for PFS;
- At the time when 155 deaths have occurred in the DDR+ population in Arms A and C combined;
- At the time when 207 deaths have occurred in the DDR+ population in Arms A and C combined.

To protect the integrity of the study and to preserve the type I error rate, a gatekeeping testing strategy is implemented and the significance levels for the interim and final efficacy analyses for each endpoint will be determined by using the Lan-DeMets procedure with an O'Brien-Fleming stopping boundary as described in Section 5.1.1.

Primary endpoint

Table 6 displays the maximum number of analyses expected for the primary endpoint and the associated efficacy and futility boundaries if the analyses are performed at the planned number of events as shown in the table. The futility boundaries are non-binding but the study may be stopped for futility if at the time of the interim analysis for PFS in the DDR+ population the associated futility boundary is crossed. If the efficacy boundary for PFS in the DDR+ population is crossed at the time of the interim analysis or at the time of the primary analysis then the primary objective of the study will have been demonstrated.

Table 6. PFS Based on BICR Assessment in the DDR+ Population (Arm A vs Arm C) - Efficacy and Futility Boundaries

	PFS in the DDR+ Population	
Analysis	**IA**	**PA**
Analysis cut-off trigger	*140 PFS events in DDR+*	*200 PFS events in DDR+*
Number of events [a] (Information fraction)	140 (70.0%)	200 (100%)
p-value (z-value) for efficacy	<0.0074 (<-2.438)	<0.0228 (<-2.000)
p-value (z-value) for futility [b]	>0.3975 (>-0.260)	Not Applicable
[a] Number of events expected under H_{11} for PFS (assuming a HR of 0.63). [b] Non-binding. IA=interim analysis , PA=primary analysis		

Since the observed number of events at the interim analysis may not be exactly equal to the planned 140 PFS events, the efficacy and futility boundaries will be updated based on the actual number of observed events using the pre-specified α-and β-spending functions. Therefore, the observed Z-test statistic at the interim analysis will be compared with the updated efficacy and futility boundaries.

If the study continues to final analysis, the p-value that will be used to declare statistical significance at the final analysis will be based on the actual number of PFS events documented at the cut-off date for the final analysis and the α already spent at the interim analysis. Therefore, if the interim analysis occurs after 140 PFS events for Arms A and C combined, and the study continues until the final analysis, the observed p-value for the comparison will have to be < 0.0228 to declare statistical significance. If the number of events in the final analysis deviates from the expected number of events, the final analysis criterion will be determined so that the overall significance level across all analyses and comparisons is maintained at 1-sided 0.025. Further details are provided in Section 7.

Based on the stopping boundaries defined above and the timing of the interim and final analyses as described above, the design has the following operating characteristics.

Table 7. Simulated Cumulative Probabilities to Stop for Efficacy or Futility at the Interim or Final PFS Analysis in the DDR+ Population (Arm A vs Arm C)

Scenario	Look	Number of PFS events	Calendar Time (months)	P(Reject H_{01})	P(Reject H_{11})
H_{01} is true (HR=1)	Interim PFS	140	24	0.0073	0.6090
	Final PFS	200	35	0.0251	0.9749
H_{11} is true (HR=0.63)	Interim PFS	140	27	0.6011	0.0066
	Final PFS	200	41	0.8990	0.1010

Simulations performed in EAST 6.4 with number of simulations = 10,000 and seed=4062018

Secondary endpoints

Two secondary endpoints, PFS in patients unselected for DDR status and OS in the DDR+ population, will be analyzed using a hierarchical testing procedure, provided the primary endpoint PFS in the DDR+ population is statistically significant favoring Arm A. Two separate α-spending functions according to Lan-DeMets (O'Brien-Fleming) independent of the one used for the primary endpoint PFS in DDR+ population will be used so that the overall 1-sided level of significance across all analyses is preserved at 0.025. The trial allows for the stopping of the study for a superior OS result, provided the primary PFS endpoint has already been shown to be statistically significant favoring Arm A.

Table 8 displays the analysis triggers for PFS in patients unselected for DDR status and the analysis triggers for OS in the DDR+ population, as well as the associated efficacy boundaries. As described in Section 5.1.1, the significance level for the analyses of these endpoints is determined by the gatekeeping procedure.

Table 8. PFS Based on BICR Assessment in Patients Unselected for DDR status and OS in the DDR+ Population (Arm A vs Arm C) - Efficacy Boundaries

	Endpoints					
	PFS in patients unselected for DDR status		OS in the DDR+ Population			
Analysis	**IA1**	**PA**	**IA1**	**IA2**	**IA3**	**PA**
Analysis cut-off trigger	*140 PFS events in DDR+*	*200 PFS events in DDR+*	*140 PFS events in DDR+*	*200 PFS events in DDR+*	*155 deaths in DDR+*	*207 deaths in DDR+*
Number of events [a] (Information fraction)	289 (70.3%)	411 (100%)	66 (31.9%)	110 (53.1%)	155 (74.9%)	207 (100%)
p-value (z-value) for efficacy	<0.0004 (<-3.352)	<0.0029 (<-2.762)	<0.00005 (<-3.891)	<0.0017 (<-2.936)	<0.0076 (<-2.428)	<0.0194 (<-2.065)

[a] Number of events expected under H_{12} for PFS (assuming a HR of 0.70) and H_{13} for OS (assuming a HR of 0.70).
IA1=interim analysis 1, IA2=interim analysis 2, IA3=interim analysis 3, PA=primary analysis

Separately for each secondary endpoint, PFS in patients unselected for DDR status and OS in the DDR+ population, the exact nominal p-values that will need to be observed to declare statistical significance at the time of these analyses will depend on the number of events that have been observed at the time of these analyses and the α already spent at the time of earlier analyses.

If PFS in patients unselected for DDR status is tested alone, independent of the testing strategy for PFS in the DDR+ population, the design concerning PFS in patients unselected for DDR status will have the following operating characteristics. These calculations are unadjusted for the pre-testing of PFS in the DDR+ population.

Table 9. Simulated Cumulative Probabilities to Stop for Efficacy at the Interim or Final PFS Analysis in Patients Unselected for DDR Status (Arm A vs Arm C)

Scenario	Look	Number of PFS events	Calendar Time (months)	P(Reject H_{02})
H_{02} is true (HR=1)	Interim PFS	324	27	0.0003
	Final PFS	446	41	0.0027
H_{12} is true (HR=0.70)	Interim PFS	288	27	0.3701
	Final PFS	410	41	0.8016

Simulations performed in EAST 6.4 with number of simulations = 10,000 and seed=50021028.
Interim and final PFS analyses calendar time expected at 27 and 41 months, respectively, under H_{11} (HR for PFS in the DDR+ population =0.63).

If OS in the DDR+ population is tested alone, independent of the testing strategy for PFS in the DDR+ population, the design concerning OS in the DDR+ population will have the

following operating characteristics. These calculations are unadjusted for the pre-testing of PFS in the DDR+ population.

Table 10. Simulated Cumulative Probabilities to Stop for Efficacy at the Interim or Final OS Analysis in the DDR+ Population (Arm A vs Arm C)

Scenario	Look	Number of OS events	Calendar Time (months)	P(Reject H_{03})
H_{03} is true (HR=1)	Interim PFS	76	27	<0.0001
	Final PFS	124	41	0.0017
	75% information fraction OS in DDR+	155	53	0.0092
	Final OS	207	79	0.0225
H_{13} is true (HR=0.70)	Interim PFS	66	27	0.0068
	Final PFS	109	41	0.1434
	75% information fraction OS in DDR+	155	61	0.4229
	Final OS	207	93	0.6960

Simulations performed in EAST 6.4 with number of simulations = 10,000 and seed=4072018.
Interim and final PFS analyses calendar time expected at 27 and 41 months, respectively, under H_{11} (HR for PFS in the DDR+ population =0.63).

The secondary endpoint, OS in patients unselected for DDR status, will be analyzed using a hierarchical testing procedure, provided the primary endpoint PFS in the DDR+ population and at least one of the previously mentioned secondary endpoints (PFS in patients unselected for DDR status and OS in DDR+ population) are statistically significant favoring Arm A. A separate α-spending function according to Lan-DeMets (O'Brien-Fleming) independent of the ones used before (ie. the three separate α-spending functions used for PFS in the DDR+ population, PFS in patients unselected for DDR status, and OS in DDR+ population) will be used so that the overall 1-sided level of significance across all analyses is preserved at 0.025.

Table 11 displays the analysis triggers for OS in patients unselected for DDR status, as well as the associated efficacy boundaries. As described in Section 5.1.1, the significance level for the analyses of this endpoint is determined by the gatekeeping procedure.

Table 11. OS in Patients Unselected for DDR Status (Arm A vs Arm C) - Efficacy Boundaries

	OS in Patients unselected for DDR status			
Analysis	**IA1**	**IA2**	**IA3**	**PA**
Analysis cut-off trigger	*140 PFS events in DDR+*	*200 PFS events in DDR+*	*155 deaths in DDR+*	*207 deaths in DDR+*
Number of events [a] (Information fraction)	135 (32.1%)	225 (53.4%)	317 (75.3%)	421 (100%)
p-value (z-value) for efficacy [b]	<0.00008 (<-3.789)	<0.0021 (<-2.856)	<0.0091 (<-2.361)	<0.0219 (<-2.016)

[a] Number of events expected under H_{14} for OS (assuming a HR of 0.75).
[b] The p-values and z-values noted for OS are those associated with the scenario when both H_{02} and H_{03} are rejected.
IA1 = interim analysis 1, IA2 = interim analysis 2, IA3=interim analysis 3, PA = primary analysis

The exact nominal p-values that will need to be observed to declare statistical significance at the time of these analyses for OS in patients unselected for DDR status will depend on which one or both of H_{02} (PFS in patients unselected for DDR status) and H_{03} (OS in DDR+ population) are rejected, the number of events that have been observed at the time of these analyses, and the α already spent at the time of earlier analyses.

If OS in patients unselected for DDR status is tested alone at 1-sided α level of 0.025, independent of the testing strategy for PFS in the DDR+ population, PFS in patients unselected for DDR status and OS in DDR+ population, the design concerning OS in patients unselected for DDR status will have the following operating characteristics. These calculations are unadjusted for any pre-testing.

Table 12. Simulated Cumulative Probabilities to Stop for Efficacy at the Interim or Final OS Analysis in Patients Unselected for DDR Status, Assuming α Level of 0.025 (Arm A vs Arm C)

Scenario	Look	Number of OS events	Calendar Time (months)	P(Reject H_{04})
H_{04} is true (HR=1)	Interim PFS	151	27	0.0001
	Final PFS	248	41	0.0022
	75% information fraction OS in DDR+	347	61	0.0096
	Final OS	450	93	0.0233
H_{14} is true (HR=0.75)	Interim PFS	135	27	0.0160
	Final PFS	223	41	0.2351
	75% information fraction OS in DDR+	317	61	0.5858
	Final OS	421	93	0.8319

Simulations performed in EAST 6.4 with number of simulations = 10,000 and seed=5022018.
The first and second interim OS analyses calendar time expected at 27 and 41 months, respectively, under H_{11} (HR for PFS in the DDR+ population =0.63).
The third interim and final OS calendar time expected at 61 and 93 months, respectively, under H_{13} (HR for OS in the DDR+ population =0.70).

5.2. General Methods

As described in Section 3.4, in this study '**treatment arm**' refers to one of the following:

- Arm A = platinum-based chemotherapy + avelumab followed by avelumab + talazoparib maintenance;

- Arm B = platinum-based chemotherapy followed by talazoparib maintenance;
- Arm C = platinum-based chemotherapy + bevacizumab followed by bevacizumab maintenance.

Endpoints will be summarized based on the analysis sets described in Table 4 by treatment arm, unless otherwise specified.

5.2.1. Data handling after the cut-off date

Data after the cut-off date may not undergo the cleaning process and will not be displayed in any listings or used for summary statistics, statistical analyses or imputations.

5.2.2. Pooling of centers

In order to provide overall estimates of treatment effects, data will be pooled across centers. The 'center' factor will not be considered in statistical models or for subgroup analyses due to the high number of participating centers in contrast to the anticipated small number of patients randomized at each center.

5.2.3. Presentation of continuous and qualitative variables

Continuous variables will be summarized using descriptive statistics ie, number of non-missing values and number of missing values [ie, n (missing)], mean, median, standard deviation (SD), minimum, maximum, and first and third quartile (Q1 and Q3).

Qualitative variables will be summarized by frequency counts and percentages. Unless otherwise specified, the calculation of proportions will include the missing category. Therefore counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of patients still present in the study at that visit, unless otherwise specified.

5.2.4. Definition of study day

Start day of study treatment is the day of the first dose of study treatment.

The study day for assessments occurring on or after the start of study treatment (eg, adverse event onset, tumor measurement) will be calculated as:

Study day = Date of the assessment/event - start of study treatment + 1.

The study day for assessments occurring prior to the first dose of study treatment (eg, baseline characteristics, medical history) will be negative and calculated as:

Study day = Date of the assessment/event - start of study treatment.

The study day will be displayed in all relevant data listings.

5.2.5. Definition of start of new anti-cancer drug therapy

Start date of new anti-cancer drug therapy is used to determine the end of the on-treatment period (see Section 5.2.7).

The start date of new anti-cancer drug therapy is the earliest start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages that is after the first dose of study treatment. When start date of anti-cancer drug therapy is missing or partially missing, the imputation rules described in Section 5.3.3.4 should be applied using only data from the 'Follow-up Cancer Therapy' eCRF pages.

5.2.6. Definition of start of new anti-cancer therapy

Start date of new anti-cancer therapy (drug, radiation, surgery) is used for censoring in efficacy analyses (see Section 6.1.1 and Section 6.2.2).

The start date of new anti-cancer therapy is the earliest date after randomization amongst the following:

- Start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages
- Start date of radiation therapy recorded in 'Concomitant Radiation Therapy', and 'Follow-up Radiation Therapy' eCRF pages with 'Treatment Intent' = 'Curative in intent'
- For neoadjuvant patients, ie, patients that answer 'Yes' to the question 'Neoadjuvant patient' in the 'Planned Neoadjuvant Therapy' eCRF page: Surgery date recorded in 'Follow-up Cancer Surgery' eCRF pages when 'Surgery Outcome' = 'Resected' or 'Partially Resected'. For patients that are NOT neoadjuvant: Surgery date recorded in 'On-Study Cancer Surgery', and 'Follow-up Cancer Surgery' eCRF pages when 'Surgery Outcome' = 'Resected' or 'Partially Resected'.

When start date of anti-cancer therapy is missing or partially missing, the imputation rules described in Section 5.3.3.4 should be applied using 'Follow-up Cancer Therapy', 'Concomitant Radiation Therapy', 'Follow-up Radiation Therapy', 'On-Study Cancer Surgery', and 'Follow-up Cancer Surgery' eCRF pages.

5.2.7. Definition of on-treatment period

Safety endpoints will be summarized based on the on-treatment period unless otherwise specified.

On-treatment period is defined as the time from the first dose of study treatment through minimum (30 days + last dose of study treatment, start day of new anti-cancer drug therapy – 1 day).

Safety data collected outside the on-treatment period as described above will be listed and flagged in listings but not summarized.

5.2.8. Standard derivations and reporting conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

Demographics and physical measurements:

- Age [years]:
  - (date of given informed consent - date of birth + 1) / 365.25
  - In case of missing day, day only: Age [years]: (year/month of given informed consent – year/month of birth)
  - In case only year of birth is given: Age [years]: (year of given informed consent - year of birth)

The integer part of the calculated age will be used for reporting purposes.

- BMI (kg/m^2) = weight (kg)/[height (m)]^2
- BSA (m^2) = ([height (cm) × weight (kg)] / 3600)$^{0.5}$

For reporting conventions, mean and median should generally be displayed one more decimal place than the raw data and standard deviation should be displayed to two more decimal places than the raw data. Percentages will be reported to one decimal place. The rounding will be performed to closest integer / first decimal using the common mid-point between the two consecutive values. Eg, 5.1 to 5.4 will be rounded to an integer of 5, and 5.5 to 5.9 will be rounded to an integer of 6.

5.2.9. Unscheduled visits

Generally, data collected at unscheduled visits will be included and analyzed for both safety and efficacy analyses in the same fashion as the data collected at scheduled visits except where otherwise noted in the sections that follow. Descriptive statistics (mean, SD, median, minimum, maximum, quartiles) by nominal visit or time point for safety endpoints such as laboratory measurements, ECGs and vital signs will include only data from scheduled visits.

5.2.10. Adequate baseline tumor assessment

Adequate baseline is defined using the following criteria:

- All baseline assessments must be within 28 days prior to and including the date of randomization.
- For patients with evidence of disease at baseline, all documented lesions must have non-missing assessments (ie, non-missing measurements for target lesions and non-missing lesions assessment status at baseline for non-target lesions).

5.2.11. Adequate post-baseline tumor assessment

For patients who undergo IDS, the IDS date will be identified as follows:

- 'Neoadjuvant Patient' question on 'Planned Neoadjuvant Therapy' eCRF page is answered 'Yes', and

- IDS date is the first surgery date recorded on the 'On-Study Cancer Surgery' eCRF page, where treatment intent is 'Curative Intent' and surgery outcome is 'Resected' or 'Partially Resected'.

For patients who do not undergo IDS, an adequate post-baseline assessment is defined as an assessment where a response of CR, PR, SD, non-CR/non-PD, or PD can be determined. Time points where the response is not evaluable (NE) or no assessment was performed will not be used for determining the censoring date.

For patients who undergo IDS, since assessments after IDS can only be NE or PD, all post-IDS tumor assessments will be considered adequate. Time points where no assessment was performed will not be used for determining the censoring date.

5.3. Methods to Manage Missing Data

5.3.1. Missing data

Unless otherwise specified, all data will be evaluated as observed, and no imputation method for missing values will be used.

In all patient data listings imputed values will be presented. In all listings imputed information will be flagged.

Missing statistics, eg when they cannot be calculated, should be presented as 'ND' or 'NA'. For example, if N=1, the measure of variability (SD) cannot be computed and should be presented as 'ND' or 'NA'.

5.3.1.1. Pharmacokinetic concentrations

Concentrations Below the Limit of Quantification

For all calculations, figures and estimation of individual pharmacokinetic parameters, all concentrations assayed as below the level of quantification (BLQ) will be set to zero. In log-linear plots these values will not be represented. The BLQ values will be excluded from calculations of geometric means and their CIs. A statement similar to 'All values reported as BLQ have been replaced with zero' should be included as a footnote to the appropriate tables and figures.

Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of median profiles, concentrations will be set to missing if one of the following cases is true:

1. A concentration has been reported as ND (ie, not done) or NS (ie, no sample);
2. A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the clinical pharmacologist.

Summary statistics will not be presented at a particular time point if more than 50% of the data are missing. For analysis of pharmacokinetic concentrations, no values will be imputed for missing data.

5.3.2. Handling of incomplete dates

5.3.2.1. Disease history

Incomplete dates for disease history (eg, initial diagnosis date, date of documented, locally advanced, inoperable or metastatic disease diagnosis, date of response or progression in prior treatment) will be imputed as follows:

- If the day is missing, it will be imputed to the 15th day of the month.
- If both day and month are missing and the year is prior to the year of the first study treatment, the month and day will be imputed as July 1st.
- If both day and month are missing and the year is same as the year of the first study treatment, the month and day will be imputed as January 1st.
- If the date is completely missing, no imputation will be performed.

5.3.2.2. Adverse events

Incomplete AE-related dates will be imputed as follows:

- If the AE onset date is missing completely, then the onset date will be replaced by the start of study treatment.
- If only the day part of the AE onset date is missing, but the month and year are equal to the start of study treatment, then the AE onset date will be replaced by the start of study treatment. For example, if the AE onset date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed AE onset date will be 15/JAN/2015.
- If both the day and month of the AE onset date are missing but the onset year is equal to the start of study treatment, then the onset date will be replaced by the start of study treatment. For example, if AE onset date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed AE onset date will be 19/NOV/2014.
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of patient's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete stop date will not be imputed. If stop date of AE is after the date of cut-off outcome of AE is ongoing at cut-off.

5.3.2.3. Prior and concomitant medications

Incomplete prior/concomitant medication dates will be imputed as follows:

- If the medication date is missing completely, then the medication date will be replaced by the start of study treatment.
- If the day of medication date is missing, but the month and year are equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. For example, if the medication start date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed medication start date will be 15/JAN/2015.
- If both the day and month of medication start date are missing but the start year is equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. For example, if the medication start date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed medication start date will be 19/NOV/2014.
- In all other cases the missing medication day or missing medication month will be replaced by 1.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of patient's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete medication stop date will not be imputed.

5.3.2.4. Exposure

No imputation will be done for first dose date. Date of last dose of study drug, if unknown or partially unknown, will be imputed as follows:

- If the last date of study drug is completely missing and there is no End of Treatment eCRF page and no death date, the patient should be considered to be ongoing and use the cut-off date for the analysis as the last dosing date
- If the last date of study drug is completely or partially missing and there is EITHER an End of Treatment eCRF page OR a death date available (within the cut-off date), then imputed last dose date is:

  = 31DECYYYY, if only Year is available and Year < Year of min (EOT date, death date)

  = Last day of the month, if both Year and Month are available and Year = Year of min (EOT date, death date) and Month < the month of min (EOT date, death date)

  = min (EOT date, death date), for all other cases.

5.3.3. Imputation rules for date of last contact and efficacy assessments

5.3.3.1. Date of last contact

The date of last contact will be derived for patients not known to have died at the analysis cut-off using the latest complete date among the following:

- All patient assessment dates (blood draws (laboratory, PK), vital signs, performance status, ECG, tumor assessments)
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation
- AE start and end dates
- Last date of contact collected on the 'Survival Follow-up' eCRF (do not use date of survival follow-up assessment unless status is 'alive')
- Study drug start and end dates
- Randomization date
- Withdrawal of consent date
- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up).

Only dates associated with actual examinations of the patient will be used in the derivation. Dates associated with a technical operation unrelated to patient status such as the date a blood sample was processed will not be used. Assessment dates after the cut-off date will not be applied to derive the last contact date.

5.3.3.2. Death date

Missing or partial death dates will be imputed based on the last contact date:

- If the date is missing it will be imputed as the day after the date of last contact
- If the day or both day and month is missing, death will be imputed to the maximum of the full (non-imputed) day after the date of last contact and the following:
  - Missing day: 1st day of the month and year of death
  - Missing day and month: January 1st of the year of death

5.3.3.3. Tumor assessments

All investigation dates (eg, X-ray, CT scan) must be completed with day, month and year.

If there are multiple scan dates associated with an evaluation, ie, radiological assessments occur over a series of days rather than the same day, the choice of date of assessment could impact the date of progression and/or date of response. If there are multiple scan dates

associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.

If one or more investigation dates for an evaluation are incomplete but other investigation dates are available, the incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the earliest of all investigation dates (eg, X-ray, CT-scan).

If all measurement dates for an evaluation have no day recorded, the 1st of the month is used.

If the month is not completed, for any of the investigations for an evaluation, the respective assessment will be considered to be at the date which is exactly between the previous and the following assessment. If both a previous and following assessments are not available, this assessment will not be used for any calculations.

5.3.3.4. Date of start of new anti-cancer therapy

Incomplete dates for start date of new anti-cancer therapy (drug therapy, radiation, surgery) will be imputed as follows and will be used for determining censoring dates for efficacy analyses and in the derivation of the end of on-treatment period. PD date below refers to PD date by investigator assessment.

- The end date of new anti-cancer therapy will be included in the imputations for start date of new anti-cancer therapy. If the end date of new anti-cancer therapy is
  - completely missing then it will be ignored in the imputations below
  - partially missing with only year (YYYY) available then the imputations below will consider 31DECYYYY as the end date of the new anti-cancer therapy
  - partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anti-cancer therapy
- For patients who have not discontinued study treatment at the analysis cut-off date, last dose of study treatment is set to the analysis cut-off date in the imputations below.
- If the start date of new anti-cancer therapy is completely or partially missing then the imputed start date of new anti-cancer therapy is derived as follows:
  - <u>Start date of new anti-cancer therapy is completely missing</u>

    Imputed start date = min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]
  - <u>Only year (YYYY) for start of anti-cancer therapy is available</u>

    IF YYYY < Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy] THEN imputed start date = 31DECYYYY;

    ELSE IF YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN imputed start date = min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

ELSE IF YYYY > Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN imputed start date = 01JANYYYY

- Both Year (YYYY) and Month (MMM) for start of anti-cancer therapy are available

IF

YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

MMM < Month of min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

THEN

imputed start date = DAY (Last day of MMM) MMM YYYY ;

ELSE IF

YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

MMM = Month of min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

THEN

imputed start date = min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]);

ELSE IF

YYYY = Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], AND

MMM > Month of min [max(PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

THEN

imputed start date = 01 MMM YYYY;

ELSE IF

YYYY < Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN

imputed start date = DAY (Last day of MMM) MMM YYYY;

ELSE IF

YYYY > Year of min [max(PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

THEN

imputed start date = 01 MMM YYYY.

5.3.3.5. PRO data

The respective scoring or user guides for NFOSI-18 and EQ-5D-5L provide missing data handling rules (FACIT, 2008; EuroQoL Group 2015). In particular, the following will be implemented.

If there are missing items on the NFOSI-18, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered, as follows:

Prorated subscale score = [Sum of item scores] × [Number of items in subscale] ÷ [Number of items answered]

When there are missing data, prorating by subscale in this way is acceptable if > 50% of the items were answered. The total score is then calculated as the sum of the un-weighted subscale scores. In addition, a total score should only be calculated if all of the component subscales have valid scores. The EQ-5D-5L questions not answered will be considered missing items and will not be utilized. For EQ-5D-5L, the entire utility score for that cycle is deemed missing if the answer to any one of the 5 dimensions is missing.

6. ANALYSES AND SUMMARIES

Refer to Section 4 for definitions of analysis sets and Section 5.2 for general methodology.

6.1. Primary Endpoint

6.1.1. Progression-free survival as assessed by BICR per RECIST v1.1 in the DDR+ population

6.1.1.1. Primary analysis

The following analyses will be based on patients in the DDR+ population in the FAS using the strata assigned at randomization. PD below refers to PD by BICR assessment.

Progression-Free Survival (PFS) is defined as the time from the date of randomization to the date of the first documentation of PD or death due to any cause, whichever occurs first.

PFS data will be censored on the date of the last adequate tumor assessment for patients who do not have an event (PD or death), for patients who start a new anti-cancer therapy prior to an event (see Section 5.2.6) or for patients with an event after 2 or more missing tumor assessments. Patients who do not have an adequate baseline tumor assessment or who do not have an adequate post-baseline tumor assessment will be censored on the date of randomization unless death occurred on or before the time of the second planned tumor assessment (ie ≤ 18 weeks after the date of randomization) in which case the death will be considered an event.

In this study antitumor activity will be assessed through radiological tumor assessments conducted at screening, at 9 and 18 weeks (±3 days) from the date of randomization and every 12 weeks thereafter until PD by BICR assessment per RECIST v1.1, regardless of initiation of subsequent anti-cancer therapies. For patients who undergo IDS during the chemotherapy period, an additional tumor assessment should be performed after surgery.

The censoring and event date options to be considered for the PFS analysis are presented in Table 13.

PFS (months) = [date of event or censoring– date of randomization +1]/30.4375

Table 13. Outcome and Event Dates for PFS Analyses

Scenario	Date of event/censoring	Outcome
No adequate baseline assessment	Date of randomization [a]	Censored [a]
PD or death – After at most one missing or inadequate post-baseline tumor assessment, OR – ≤ 18 weeks after the date of randomization	Date of PD or death	Event
PD or death – After 2 or more missing or inadequate post-baseline tumor assessments	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
No PD and no death	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
Treatment discontinuation due to 'Disease progression' without documented progression	Not applicable	Information is ignored. Outcome is derived based on documented progression only.
New anti-cancer therapy given	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored

[a] However if the patient dies ≤18 weeks after the date of randomization the death is an event with date on death date

[b] If there are no adequate post-baseline assessments prior to PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria were met the censoring will be on the date of randomization.

The primary efficacy analysis will compare the PFS time based on BICR assessment between Arm A and Arm C in the DDR+ population, and will be performed using a 1-sided stratified log-rank test as described in Section 5.1.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=control arm, 1= experimental arm) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

In order to account for the group sequential design in this study, the repeated CI (RCI) method (Jennison and Turnbull, 2000), will be used to construct the 2-sided RCIs for the hazard ratio at the interim and the final analyses of PFS. In addition, the unadjusted 95% CIs for the hazard ratio will also be reported at the interim and the final analyses for PFS.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median PFS time with 2-sided 95% CIs. In particular, the PFS rate at 9 and 18 weeks and at 8, 12, 18, 24, 30, 36, 42, 48, 54, and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with each event type (PD or death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 14 following the hierarchy shown.

Table 14. PFS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment	No adequate baseline assessment
2	Start of new anti-cancer therapy	Start of new anti-cancer therapy
3	Event after 2 or more missing or inadequate post-baseline tumor assessments/date of randomization	Event after 2 or more missing assessments [a]
4	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Subject refused further follow-up]	Withdrawal of consent
5	No event and lost to follow-up in any disposition page	Lost to follow-up
6	No event and [EOS present OR disposition page for any epoch after screening says patient will not continue into any subsequent phase of the study] and no adequate post-baseline tumor assessment	No adequate post-baseline tumor assessment
7	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

[a] 2 or more missing or inadequate post-baseline tumor assessments.

The PFS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Time of Follow-Up for PFS

A plot will be generated to compare planned and actual relative day of tumor assessments by treatment arm. A Kaplan-Meier plot for PFS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the PFS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median time of follow-up for PFS with 2-sided 95% CIs. In particular, the rate at 9 and 18 weeks and at 8, 12, 18, 24, 30, 36, 42, 48, 54, and 60 months will be estimated with corresponding 2-sided 95% CIs.

6.1.1.2. Analysis in Arm B versus Arm C

The analyses described in Section 6.1.1.1 will be repeated for the Arm B vs Arm C assessment; however this comparison is not part of the formal testing strategy for the study as described in Section 5.1 and, therefore, RCI will not be applicable and p-values reported will be purely descriptive.

The probability of the HR (Arm B vs Arm C) in the DDR+ population ($HR_{PFS+(BvsC)}$) being less than 1 will be calculated assuming asymptotic normality of $\log(HR_{PFS+(BvsC)})$ and a non-informative prior. Proof of efficacy for Arm B compared to Arm C will have been demonstrated if both of the following two criteria are satisfied:

- Clinical relevance: Estimated $HR_{PFS+(BvsC)} \leq 0.792$.
- Statistical significance: Probability (True $HR_{PFS+(BvsC)} < 1 \mid \text{data}) \geq 0.80$.

6.2. Secondary Endpoint(s)

6.2.1. Safety endpoints

Refer to Section 6.6.

6.2.2. Efficacy endpoints

The following analyses will be based on the FAS by treatment arm unless otherwise specified.

PFS by investigator assessment will be analyzed as a secondary endpoint using the same methodology that is described in Section 6.1.1.1 now referring to PD by investigator assessment instead of PD by BICR assessment; RCI will not be calculated.

6.2.2.1. Progression-free survival as assessed by BICR per RECIST v1.1 in patients unselected for DDR status

The analyses described in Section 6.1.1.1 will be repeated for patients unselected for DDR status:

- Arm A vs Arm C comparison will use a 1-sided stratified log-rank test as described in Section 5.1, and

- Arm B vs Arm C assessment will be performed; RCI will not be reported and p-values will be nominal.

6.2.2.2. Overall survival in the DDR+ population and in patients unselected for DDR status

The following analyses will be based on the FAS using the strata assigned at randomization.

Overall survival (OS) is defined as the time from the date of randomization to the date of death due to any cause. Patients last known to be alive will be censored at date of last contact.

OS (months) = [date of death or censoring– date of randomization +1]/30.4375

The analysis of OS will compare the OS time between Arm A and Arm C for patients in the DDR+ population and for patients unselected for DDR status, and will be performed using a 1-sided stratified log-rank test as described in Section 5.1. OS time between Arm B and Arm C will also be assessed (p-values reported will be purely descriptive) for patients in the DDR+ population and for patients unselected for DDR status.

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=control arm, 1= experimental arm) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

In order to account for the group sequential design in this study, the repeated CI (RCI) method (Jennison and Turnbull, 2000), will be used to construct the 2-sided RCIs for the hazard ratio at the interim and the final analyses of OS. In addition, the unadjusted 95% CIs for the hazard ratio will also be reported at the interim and the final analyses for OS.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median OS time with 2-sided 95% CIs. In particular, the OS rate at 9 and 18 weeks and at 8, 12, 18, 24, 30, 36, 42, 48, 54, 60, 78, 96, 120, and 144 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with an event (death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 15.

Table 15. OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Subject refused further follow-up]	Withdrawal of consent
2	No event and [lost to follow-up in any disposition page OR data cut-off date – last contact date > 14 weeks]	Lost to follow-up
3	No event and none of the conditions in the prior hierarchy are met	Alive

The OS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Time of Follow-Up for OS

A Kaplan-Meier plot for OS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the OS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median time of follow-up for OS with 2-sided 95% CIs. In particular, the rate at 9 and 18 weeks and at 8, 12, 18, 24, 30, 36, 42, 48, 54, 60, 78, 96, 120, and 144 months will be estimated with corresponding 2-sided 95% CIs.

6.2.2.3. Sensitivity analyses for progression-free survival

All sensitivity analyses described below will be performed for PFS based on BICR in the DDR+ population and separately for PFS based on BICR in patients unselected for DDR status.

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results. These analyses are regarded as purely exploratory. The sensitivity analyses will repeat the primary analysis (p-value, HR and 95% CIs) described in Section 6.1.1.1 with the modifications below:

- PFS based on BICR assessment and counting all PD and deaths as events regardless of missing assessments or timing of the event
- PFS based on BICR assessment on the PP analysis set for PFS
- PFS based on BICR assessment using an unstratified analysis
- PFS based on BICR assessment using strata recorded on eCRF instead of those entered in IRT

- PFS based on BICR assessment stratified by resection status per eCRF and BRCA1/2 mutational status as determined by foundation medicine
- PFS based on BICR assessment modifying the censoring rules in Table 13 to consider all deaths as events
- PFS based on BICR assessment modifying the censoring rules in Table 13 with initiation of subsequent anti-cancer therapies not used as a censoring reason.

Methods for evaluating the validity of model assumptions

The proportional hazards assumption will be checked visually by plotting log(-log(PFS)) versus log(time) within each randomization stratum.

Schoenfeld residuals for the stratified Cox proportional regression model will be plotted to investigate graphically violations from the proportional hazards (PH) assumption; a non-zero slope is evidence of departure from PH. The PH assumption will be formally tested using Schoenfeld's residual test (Schoenfeld, 1980; Therneau & Grambsch, 2000). Large departures from PH will be evidenced by a p-value <0.05.

If these show large departures from proportional hazards, then PFS by BICR assessment will also be analyzed based on restricted mean survival time (RMST) differences (Zhang, 2013).

Restricted Mean Survival Time (RMST)

The hazard ratio estimate from the Cox proportional hazard model is routinely used to empirically quantify the between-arm difference under the assumption that the ratio of the two hazard functions is constant over time. When this assumption is plausible, such a ratio estimate may capture the relative difference between two survival curves. However, the clinical meaning of such a ratio estimate is difficult, if not impossible, to interpret when the underlying PH assumption is violated (ie, the hazard ratio is not constant over time).

The RMST is a robust and clinically interpretable summary measure of the survival time distribution. Unlike median survival time, it is estimable even under heavy censoring. There is a considerable body of methodological research (eg, Royston and Parmar, 2011; Uno et al., 2014; Zhang, 2013) about the use of RMST to estimate treatment effects as an alternative to the hazard ratio approach.

The RMST methodology is applicable independently of the PH assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis results. However, when large departures from the PH assumption are observed, the log-rank test is underpowered to detect differences between the survival distributions for the treatment arms, and a test of the difference between the RMST for the experimental arm and the control arm may be more appropriate to determine superiority of the experimental arm compared to the control arm with respect to the time-to-event endpoint.

In particular, as it pertains to the **cut-off point (τ)** to evaluate the RMST, it is noted that the cut-off point should not exceed the minimum of the largest observed time for both treatment arms so that the RMST of all treatment arms being evaluated can be adequately estimated and comparison between treatments is feasible; τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval. The RMST up to time τ can then be interpreted as the expected survival time restricted to the common follow-up time τ among all patients. The selection of τ should ensure that the RMST evaluation will not go beyond the maximum time point where the evaluation can be performed while also taking into account a large period of time that is expected to provide a meaningful assessment of treatment effect. To avoid arbitrary selection of the common cut-off τ for both treatment arms, three sets of analyses will be performed:

- τ_1 = minimum of (largest observed PFS time for the experimental arm, largest observed PFS time for the control arm).
- τ_2 = minimum of (largest PFS event time for the experimental arm, largest PFS event time for the control arm).
- τ_3 = midpoint between τ_1 and τ_2.

The treatment effect between the experimental arm and the control arm will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and 1-sided p-value will be generated.

BICR vs Investigator Assessment

A summary of the BICR assessment versus investigator assessment will be provided including numbers of concordant and discordant assessments as well as the number of cases where PFS event was assessed at different timepoints based on BICR and investigator assessments.

Table 16 outlines the possible outcomes by investigator and BICR (Amit et al. 2011).

Table 16. Possible Outcomes for Investigator vs BICR

		BICR	
		Event	**No Event**
Investigator	**Event**	a = a1 + a2 + a3	b
	No Event	c	d

a1: number of agreements on timing and occurrence of event;

a2: number of times agreement on event but INV declares event later than BICR;

a3: number of times agreement on event but INV declares event earlier than BICR;

N= a+b+c+d.

The timing agreement of event is defined as a window of ± 7 days.

The following measure of discordance will be calculated for each treatment arm:

- Total Event Discrepancy Rate: (b+c) / N
- Early Discrepancy Rate (EDR): (a3+b) / (a+b)
- Late Discrepancy Rate (LDR): (a2+c) / (a2+a3+b+c)
- Overall Discrepancy Rate: (a2+a3+b+c) / N

The EDR represents the positive predictive value of investigator assessment and quantifies the frequency with which the investigator declares PFS event earlier than BICR within each treatment arm as a proportion of the total number of investigator assessed events.

The LDR quantifies the frequency with which the investigator declares PFS event later than BICR as a proportion of the total number of discrepancies within the treatment arm.

Discordance metrics are calculated for each treatment arm and, for each metric, the difference in discordance between the experimental and control arms is used to evaluate potential bias. If the discordance is similar across the treatment arms then this suggests the absence of evaluation bias favoring a particular treatment arm. A negative differential discordance for EDR and/or a positive differential discordance for LDR may be indicative of investigator evaluation bias in favor of the experimental arm (Amit et al, 2011).

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors

See subgroups as defined in Section 6.3.3.

Multivariable Cox regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The Cox's Proportional Hazard model is defined as:

$$h(t) = h(0;t)\, e^{Xb}$$

where h(0;t) defines the baseline hazard function and X defines the vector of explanatory variables and b the unknown vector of regression parameters.

In the stepwise selection procedure, variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). This analysis will be performed using the stepwise selection method in SAS (Proc PHREG). Once this procedure stops, the factor 'treatment arm' will be added to the last selected model in order to evaluate the effect of

treatment on PFS time when adjusted for the selected explanatory variables. The hazard ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% CIs. No interactions will be considered. Post-baseline factors will not be considered for the model.

6.2.2.4. Sensitivity analyses for overall survival

All sensitivity analyses described below will be performed for OS in the DDR+ population and separately for OS in patients unselected for DDR status.

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results for OS. These analyses are regarded as purely exploratory. The sensitivity analyses will repeat the primary analysis (p-value, HR and 95% CIs) described in Section 6.2.2.2 with the modifications below:

- PP analysis set;
- unstratified;
- using strata recorded on eCRF instead of that entered in IRT
- stratified by resection status per eCRF and BRCA1/2 mutational status as determined by foundation medicine

Methods for evaluating the validity of model assumptions

The same methodology described in Section 6.2.2.3 for PFS will be used for OS.

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors

The same methodology described in Section 6.2.2.3 for PFS will be used for OS.

6.2.2.5. PFS2

The analysis of PFS2 will be performed for the DDR+ population and for patients unselected for DDR status based on FAS.

PFS2 is defined as time from the date of randomization to the start of second subsequent treatment after first PD by Investigator assessment, or death from any cause, whichever occurs first.

PFS2 (months) = [date of event or censoring – date of randomization +1]/30.4375

A patient will be considered to have an event if

1) the patient had objective PD on or prior to start of next-line anti-cancer treatment, AND started a second subsequent anti-cancer treatment

OR

2) the patient died.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median PFS2 time with 2-sided 95% CIs. In particular, the PFS2 rate at 12, 24, 36, 48, 60, and 80 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

The censoring and event date options to be considered for PFS2, each corresponding censoring reason and its hierarchy are presented in Table 17 . Frequency (number and percentage) of patients with an event (second subsequent anti-cancer treatment or death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to Table 17 following the hierarchy shown.

Table 17. Outcome, Event Dates, Censoring Reasons and Hierarchy for PFS2 Analyses

Scenario	**Date of event/ censoring**	**Outcome/ Censoring reason/ Censoring hierarchy**
(No PD[a]) and (no death)	Date of last adequate tumor assessment[b] documenting no PD	Censored/ No PD/ 1
(No PD[a]) and death	Date of death	Event (Death)
(PD[a] date > NTX1[c] start date) and (no death)	Start date of NTX1[c]	Censored/ Start of new anti-cancer treatment before PD/ 2
(PD[a] date > NTX1[c] start date) and death	Date of death	Event (Death)
(PD[a] date ≤ NTX1[c] start date) and (no death)		
• If NTX2[d] start date is non-missing	Start date of NTX2[d]	Event (Start of second subsequent anti-cancer treatment)
• Else if [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Subject refused further follow-up]	(Withdrawal of consent date) or (EOS visit date where subject refusal of further follow-up is recorded)	Censored/ Withdrawal of consent/ 3
• Else if [lost to follow-up in any disposition page]	Last contact date	Censored/ Lost to follow-up/ 4
• Else if no prior conditions are met	Last contact date	Censored/ Ongoing without PFS2 event/

Scenario	Date of event/ censoring	Outcome/ Censoring reason/ Censoring hierarchy
		5
(PD[a] and no NTX1[c]) and (no death)	Last contact date	Censored/ Ongoing without PFS2 event/ 6
• (PD[a] date ≤ NTX1[c] start date) and death		
• If NTX2[d] start date is non-missing	Start date of NTX2[d]	Event (Start of second subsequent anti-cancer treatment)
• Else if the prior condition is not met	Date of death	Event (Death)

[a] PD is the first PD by investigator assessment per RECIST v1.1, without considering any censoring rules.

[b] If there are no adequate post-baseline assessments, then the censoring date is the date of randomization. If patient has initiated next-line anti-cancer treatment, the last adequate post-baseline assessment on or prior to start date of next-line anti-cancer treatment will be considered.

[c] NTX1 is the first new anti-cancer treatment.

[d] NTX2 is the second new anti-cancer treatment.

The PFS2 time or censoring time and the reasons for censoring will also be presented in a patient listing.

6.2.2.6. PFS by GCIG criteria based on investigator assessment

The analysis of PFS by GCIG criteria will be performed for the DDR+ population and for patients unselected for DDR status based on FAS.

PFS from the date of randomization by GCIG criteria will be assessed in this study incorporating both RECIST v1.1 and CA-125 (Rustin, G et.al.) (see Appendix 6 of the protocol) based on Investigator assessment.

PFS by GCIG criteria will be censored if both PFS per RECIST v1.1 and PFS per CA-125 are censored, the date of censoring will be the latest of the two censoring dates.

CA-125 data will be censored on the date of the last CA-125 assessment for patients who start new anti-cancer therapy prior to an event, or for patients with an event after 2 or more missing CA-125 assessments. Patients who do not have an adequate baseline CA-125 assessment or who do not have an adequate post-baseline CA-125 assessment will be censored on the day of randomization with a duration of 1 day.

PFS by GCIG criteria (months) = [date of event or censoring – date of randomization +1]/30.4375

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median PFS by GCIG criteria time with 2-sided 95% CIs. In particular, the PFS by GCIG criteria rate at 9 and 18 weeks and at 8, 12, 24, 30, 36, 42, 48, 54 and 60 months will be estimated with corresponding 2-sided 95% CIs.

The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of patients with each event type (PD by Investigator assessment per RECIST v1.1 or PD by CA-125 or death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the hierarchy shown in Table 18.

Table 18. Censoring Reasons and Hierarchy for PFS by GCIG Criteria Based on Investigator Assessment

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment per RECIST v1.1 and no adequate CA-125 baseline assessment	No adequate baseline assessment
2	Start of new anti-cancer therapy	Start of new anti-cancer therapy
3	Event by Investigator assessment per RECIST v1.1 after 2 or more missing or inadequate post-baseline tumor assessments/date of randomization	Event per RECIST v1.1 after 2 or more missing assessments [a]
4	Event per CA-125 after 2 or more (ie, > 12 weeks) missing post-baseline assessments	Event per CA-125 after 2 or more missing assessments
5	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Subject refused further follow-up]	Withdrawal of consent
6	No event and lost to follow-up in any disposition page	Lost to follow-up
7	No event and [EOS present OR disposition page for any epoch after screening says patient will not continue into any subsequent phase of the study] and no adequate post-baseline tumor assessment and no post-baseline CA-125 assessment	No adequate post-baseline tumor assessment and no adequate post-baseline CA-125 assessment
8	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

[a] 2 or more missing or inadequate post-baseline tumor assessments

The PFS by GCIG criteria time or censoring time and the reasons for censoring will also be presented in a patient listing.

6.2.3. Pharmacokinetic endpoints

The following pharmacokinetic analyses will be based on the PK analyses set by treatment arm for Arms A (avelumab and talazoparib) and B (talazoparib only).

C_{trough} and C_{max} for avelumab and C_{trough} for talazoparib will be reported. Dose normalized C_{trough} for talazoparib [$C_{trough}(dn)$] will be reported as appropriate.

Pharmacokinetic parameters for avelumab and talazoparib will be taken from observed values as described in Section 3.2.4.

Presentation of pharmacokinetic data will include:

- One set of tables by analyte will be generated for plasma concentrations/PK parameters C_{trough} and C_{max} for avelumab and C_{trough} and $C_{trough}(dn)$ (if needed) for talazoparib.
- Parameters will be listed and summarized by treatment arm, cycle, day and nominal time using descriptive statistics (n, mean, SD, %CV, median, minimum, maximum, geometric mean and its associated %CV, and 95% CI). Data from patients who undergo intrapatient dose escalation or reduction for talazoparib or avelumab will be excluded from summary statistics for C_{trough} and C_{max} from the time the dose escalation or reduction occurs. For talazoparib, $C_{trough}(dn)$ will be presented only if the patient undergoes intrapatient dose escalation or reduction.
- C_{max} and C_{trough} for avelumab and C_{trough} for talazoparib will be plotted for each treatment arm using a box-whisker plot by cycle and day within cycle in order to assess the attainment of steady-state. Data from patients who undergo intrapatient dose escalation or reduction will be excluded from the time the dose escalation or reduction occurs. Individual data points, the geometric mean and the median of the parameter for each analyte in each treatment arm will be overlaid on the box plots. If a treatment arm has limited evaluable PK data (n<4), matchstick plots showing changes in the parameter in individual patients will then be generated. The geometric mean of the parameter in each treatment will be overlaid in the plots.

6.2.4. Population pharmacokinetic endpoints

Pharmacokinetic and pharmacodynamic data from this study may be analyzed using modeling approaches and may also be pooled with data from other studies to investigate any association between avelumab and talazoparib exposure and biomarkers or significant safety/efficacy endpoints. The results of these analyses, if performed, may be reported separately.

6.2.5. Biomarker endpoints

Secondary endpoints in the study are candidate predictive biomarkers in tumor tissue including PD-L1 expression, tumor mutational burden, genomic scarring and mutations in genes associated with DDR at baseline.

Biomarker data will be analyzed based on the biomarker analysis sets as defined in Section 4.3.3, by treatment arm.

If a patient has more than one result at a visit for a specific biomarker analyte, then:

- For continuous data, the duplicate results will be averaged, and the average used in the analysis;

- For non-continuous data (eg, identified genes), the study team will select the record appropriate for analysis. A flag will be added to the data sets indicating which record was selected for analysis.

For PD-L1 expression, patients will be classified as positive and negative according to scoring algorithms and cut-offs established from internal or external sources. Patients whose status cannot be determined are not considered to have screening biomarker assessment per the biomarker analysis set definition, and therefore will be excluded.

For continuous measurement biomarker results, summary statistics (eg, the mean, standard deviation, median, percent of coefficient of variation, and minimum/maximum levels) will be determined at baseline and on-treatment/end of treatment time points, as appropriate.

Change from baseline measurements will be provided, as appropriate.

For discrete measurement biomarker results (eg, tumor marker status), frequencies and percentages of categorical biomarker measures will be determined at baseline and on-treatment/end of treatment time points. Shift tables may also be provided.

Biomarker subgroups as defined above will be used to perform subgroup analyses for efficacy endpoints (PFS by BICR assessment, OS) using the methodology outlined in Section 6.3.3. In addition, the hazard ratio for the biomarker subgroup level comparisons and the unadjusted 95% CIs for the hazard ratio will be reported for each treatment arm.

CCI [redacted]

6.2.6. Endpoints for immunogenicity data of avelumab

All analyses described below are performed in Arm A only.

Blood samples for avelumab immunogenicity testing will be collected predose on Day 1 of chemotherapy Cycles 1-4 during the chemotherapy period. In the maintenance period, blood samples will be collected predose on Days 1 and 29 of maintenance Cycle 1. Thereafter, immunogenicity samples will be collected predose on Day 1 of maintenance cycles 2, 4, 6, 10, and at the end of treatment.

Samples positive for ADA will be analyzed for titer and may be analyzed for nAb.

Patients will be characterized into different ADA categories based on the criteria defined in Table 19.

Table 19. Patients Characterized Based on Anti-Drug Antibody Results (ADA Status)

Category	Definition	Subjects at Risk (Denominator for Incidence)
ADA never-positive	No positive ADA results at any time point; ADA-negative patients (titer < cutpoint)	Number of patients with at least one valid ADA result at any time point
ADA ever-positive	At least one positive ADA result at any time point; ADA-positive patients (titer ≥ cutpoint)	Number of patients with at least one valid ADA result at any time point
Baseline ADA positive	A positive ADA result at baseline	Number of patients with valid baseline ADA result
Treatment-boosted ADA	A positive ADA result at baseline and the titer ≥ 8×baseline titer at least once after treatment with avelumab	Number of patients with valid baseline ADA results and at least one valid post-baseline ADA result
Treatment-induced ADA	Patient is ADA-negative at baseline and has at least one positive post-baseline ADA result; or if patient does not have a baseline sample, the patient has at least one positive past-baseline ADA result	Number of patients with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Transient ADA response	If patients with treatment-induced ADA have (a single positive ADA result or duration between first and last positive result <16 weeks) and ADA result at the last assessment is not positive.	Number of patients with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Persistent ADA response	If patients with treatment-induced ADA have duration between first and last positive ADA result ≥16 weeks or a positive ADA result at the last assessment	Number of patients with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)

ADA: anti-drug antibody, NR = not reportable.

Patients will be characterized into different nAb categories based on the criteria in Table 20. For nAb, treatment-boosted is not applicable since no titer result is available.

Table 20. Patients Characterized Based on Neutralizing Antibody Results (nAb Status)

Category	Definition	Subjects at Risk (Denominator for Incidence)
nAb never-positive	No positive nAb results at any time point	Number of patients with at least one valid ADA result at any time point
nAb ever-positive	At least one positive nAb result at any time point	Number of patients with at least one valid ADA result at any time point
Baseline nAb positive	A positive nAb result at baseline	Number of patients with valid baseline ADA result
Treatment-induced nAb	Patient is not nAb positive at baseline and has at least one positive post-baseline nAb result; or if patient does not have a baseline sample, the patient has at least one positive past-baseline ADA result	Number of patients with at least one valid post-baseline ADA result and without positive baseline nAb result (including missing, NR)
Transient nAb response	If patients with treatment-induced nAb have (a single positive nAb result or duration between first and last positive result <16 weeks) and nAb result at the last assessment is not positive.	Number of patients with at least one ADA valid post-baseline result and without positive baseline nAb result (including missing, NR)
Persistent nAb response	If patients with treatment-induced nAb have duration between first and last positive nAb result ≥16 weeks or a positive nAb result at the last assessment	Number of patients with at least one valid post-baseline ADA result and without positive baseline nAb result (including missing, NR)

ADA = antidrug antibody, nAb = neutralizing antibody, NR = no result.

The number and percentage of patients in each ADA and nAb category will be summarized.

6.2.6.1. Time to and Duration of ADA and nAb response

The ADA and nAb analyses described below will include patients with treatment-induced ADA or nAb, respectively.

Time (weeks) to ADA response is defined as:

(Date of first positive ADA result – date of first dose of avelumab + 1)/7.

Time to ADA response will be summarized using simple descriptive statistics (mean, SD, median, min, max. Q1, Q3).

Duration (weeks) of ADA response is defined as:

(Date of last positive ADA result – date of first positive ADA result + 1)/7.

Duration of ADA response will be censored if:

- the last ADA assessment is positive AND patient is ongoing treatment with avelumab, or

- the last ADA assessment is positive AND patient discontinued treatment with avelumab AND the last planned ADA assessment (post-treatment 30-day follow-up) is after the cut-off date.

Time to nAb response and duration of nAb response are defined similarly based on first and last positive nAb result.

Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including the median ADA response time with 2-sided 95% CIs. ADA response rates at different timepoints will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Duration of ADA response will be displayed graphically and analyzed using Kaplan-Meier methodology. If the number of patients with ADA response is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided

As data permit, the analyses described above will be repeated for patients with treatment-induced nAb.

6.2.6.2. ADA titer

For patients who are ADA ever positive, the maximum observed ADA titer for a patient will be summarized, overall and by ADA subcategories (baseline ADA positive, treatment-boosted ADA, treatment-induced ADA, transient ADA response, persistent ADA response) of patients having each discrete maximum titer value will be tabulated. The denominator to calculate the percentages will be the total number of patients in the associated ADA subcategory.

For patients with treatment-induced ADA, a cross tabulation of duration of ADA response and maximum ADA titer will be provided. The following categories for duration of ADA response will be used: ≤1, >1 to ≤3, >3 to ≤5, >5 to ≤7, >7 to ≤13, >13 to ≤16, >16 to ≤25, >25 weeks. In this categorization, the censoring in duration of ADA response is ignored.

6.2.6.3. Analysis of PK, safety and efficacy by immunogenicity status

The following ADA and nAb status will be used for the analyses described below.

<u>ADA</u>

- ADA ever-positive versus ADA never-positive

- ADA: treatment-induced ADA versus ADA never-positive or baseline ADA positive

nAb

- nAb ever-positive versus nAb never-positive
- nAb: treatment-induced nAb versus nAb never-positive or baseline nAb positive

Data listings will include immunogenicity data together with relevant PK, safety and efficacy data.

PK parameters and immunogenicity status

The following analyses will include patients in both the immunogenicity analysis set and in the PK parameter analysis set. The PK endpoints pertinent to the immunogenicity analyses are C_{trough} and C_{max}.

During the chemotherapy period, blood samples for avelumab PK will be collected within 2 hours prior to and at the end of infusion (within 10 minutes after the avelumab infusion ends) on Day 1 of Chemotherapy Cycles 1-4. During the maintenance period, PK samples will be collected within 1 hour prior to taking talazoparib dose and the post-dose avelumab PK sample should be taken within 10 minutes after the avelumab infusion ends on Days 1 and 29 of the first maintenance cycle. Thereafter, PK samples will be collected predose within 1 hour prior to taking talazoparib dose and the post-dose avelumab PK samples should be taken within 10 minutes after the avelumab infusion ends on Day 1 of maintenance cycles 2, 4, 6, 10 and at the end of treatment.

C_{trough} and C_{max} will be summarized descriptively (n, mean, SD, CV, median, minimum, maximum, geometric mean, its associated CV, and 95% CI) by nominal time and ADA status. Linear-linear plots of mean and median for C_{trough} and C_{max} over nominal time and by ADA status will be presented.

Among patients with treatment-induced ADA, analyses will be conducted to assess whether there are any changes in C_{trough} and/or C_{max} after the first positive ADA assessment relative to before the first positive ADA assessment. To be included in this analysis, patients must have the same PK parameter available both before and after the first positive ADA assessment. Relative PK day will be calculated as:

(PK assessment nominal day) – (first positive ADA assessment nominal day).

Nominal day is the protocol scheduled timing for an assessment. For example, if C_{trough} is collected on Day 1 of Cycle 2 in the chemotherapy period and the first positive ADA result is observed on Day 1 of Cycle 3 in the chemotherapy period, then the relative PK day for this Cycle 2 day 1 C_{trough} is -21. Linear-linear plots of mean and median C_{trough} and C_{max} over relative PK day will be presented.

As data permit, the analyses described above will be repeated for nAb.

Safety and immunogenicity status

The following analyses will include patients in the immunogenicity analysis set.

The frequency (number and percentage) of patients with each of the following will be presented by ADA status.

- TEAEs, by SOC and PT
- TEAEs leading to dose reduction of avelumab, by SOC and PT
- TEAEs leading to discontinuation of avelumab, by SOC and PT
- TEAEs leading to discontinuation of study treatment by SOC and PT
- Grade ≥ 3 TEAEs, by SOC and PT
- SAEs, by SOC and PT
- IRRs, by PT

For patients who had at least one IRR and have treatment-induced ADA, time related to first onset of an IRR (infusion 1, infusion 2, infusion 3, infusion 4, or later) will be summarized taking into account whether the IRR occurred on or after the first ADA positive assessment or whether the IRR occurred before the first ADA positive assessment.

As data permit, the analyses described above will be repeated for nAb.

Efficacy and immunogenicity status

For the ADA ever-positive patients, a listing will be prepared with patient ID, start and stop of avelumab treatment, date of first positive ADA result, time to ADA response, duration of ADA response, date of last ADA positive result, PFS time or censoring time and reason for censoring, and OS time or censoring time and reason for censoring. If applicable, date of first positive nAb result, time to nAb response, duration of nAb response, and date of last nAb positive result will also be presented. PFS will be presented based on BICR assessment and based on Investigator assessment.

For the ADA ever-positive patients, the percent change from baseline in target lesions as well as the first occurrence of a new lesion and patient off avelumab treatment will be displayed against time point (weeks) in a line plot. Additional symbols will indicate the first and last ADA positive result and, if applicable, the first and last nAb positive result. Plot will be presented separately based on BICR assessment and based on Investigator assessment.

CCI

CCI

CCI

CCI

CCI

CCI [redacted]

[redacted]

6.3.2. PRO endpoints

All PRO analyses will be based on the FAS, by treatment arm and will include all the assessments per schedule of assessment from baseline to the last PRO assessment, unless otherwise specified.

The primary PRO analysis is designed to determine the effect of Arm A compared to Arm C on the time to deterioration (TTD) of the DRS-P in the DDR+ population, from randomization up to EOT (not including EOT). TTD is defined as the time from randomization to deterioration where deterioration is defined as a ≥3-point decrease, maintained for 2 consecutive assessments, on the DRS-P (9 items) for physical symptoms or concerns of disease.

6.3.2.1. Scoring procedure

The NFOSI-18 and EQ-5D-5L will be scored according to their respective scoring or user guides (FACIT, 2008; EuroQoL Group 2015). For the EQ-5D-5L, the index scores will be

calculated using the published weights (tariffs) for the United Kingdom. Specific country weights may be applied for country specific analyses as needed.

6.3.2.2. Instrument completion rates

For each treatment arm and at each time point, the number and percentage of patients who complete the NFOSI-18 and EQ-5D-5L will be summarized, as will the reasons for non-completion of these measures. An instrument is considered complete if at least one item was answered by the patient.

6.3.2.3. Descriptive summaries over time

Absolute scores and change from baseline for the NFOSI-18 total scores, subscales and the single item "I am bothered by side effects of treatment", as well as the EQ-5D-5L index and EQ-VAS will be summarized (as described in Section 5.2.3) at each time point by treatment arm. Line charts depicting the means and mean changes from baseline along with SE error bars over time will be provided for each scale by treatment arm.

For the EQ-5D-5L health status, the proportions of patients reported having "none", "slight", "moderate", "severe", or "extreme/unable" problems at each time point will be presented by treatment arm.

6.3.2.4. Time-to-event endpoints

Time to deterioration (TTD) of DRS-P

The TTD for DRS-P will be calculated from the date of randomization to the date of the first report of a score of ≥3-point decrease of the 2 consecutive assessments used to identify the event. A decline in DRS-P score represents an increase in physical symptoms or concerns of the disease. Patients who do not have a TTD event will be censored on the last date when they have a DRS-P score.

The primary TTD analysis will include all assessments from baseline up to EOT (not including EOT) to assess the level of treatment side effect bother on patients during the treatment period. The sensitivity TTD analysis will include all assessments from baseline, post-baseline, EOT and until 3 years from patient' enrollment to assess the level of treatment side effect bother on patients from the next line of therapy.

Three additional sensitivity analyses for TTD will be conducted: based on a ≥2-point, 4-point and 5-point decrease from randomization in the DRS-P, maintained for 2 consecutive for up to EOT (not including EOT).

TTD will be compared between treatment arms using a 1-sided log-rank test stratified by randomization stratification factors. No adjustment for multiplicity will be conducted.

The treatment effect, as measured by the hazard ratio, will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie, for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm

(0=reference arm, 1= non-reference arm) and β is the unknown regression parameter. Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG). The unadjusted 95% CIs for the hazard ratio will be reported.

Kaplan-Meier estimates will be presented by treatment arm together with a summary of associated statistics including the median TTD time with 2-sided 95% CIs. In particular, the TTD rate at 9 and 18 weeks and at 8, 12, 18, 24, 30, 36, 42, 48, 54 and 60 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

The TTD cut-off for the DRS-P will be also examined using an anchor-based method with the PGI-S and PGI-C items and the cumulative distribution function as part of a supplemental analysis, and reported separately from the clinical study report.

Time to development of significant side effect bother

TTB is defined as the time from the randomization date to the first report of a score of ≥2 for at least two consecutive assessments on the side effect bother item. Patients will be censored at the last time when they completed the side effect bother item if they have not had a TTB event.

The primary TTB analysis will include all assessments from baseline up to EOT (not including EOT) to assess the level of treatment side effect bother on patients during the treatment period. The sensitivity TTB analysis will include all assessments from baseline, post-baseline, EOT and until 3 years from patient' enrollment to assess the level of treatment side effect bother on patients from the next line of therapy.

TTB will be compared between treatment arms using a 1-sided log-rank test stratified by randomization strata. No adjustment for multiplicity will be conducted.

The treatment effect, as measured by the hazard ratio, will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), ie, for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=reference arm, 1= non-reference arm) and β is the unknown regression parameter. Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG). The unadjusted 95% CIs for the hazard ratio will be reported.

Kaplan-Meier estimates will be presented by treatment arm together with a summary of associated statistics including the median TTB time with 2-sided 95% CIs. In particular, the TTB rate at 9 and 18 weeks and at 8, 12, 18, 24, 30, 36, 42, 48, 54 and 60 months will be

estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

6.3.2.5. Continuous endpoints

Mixed-effects longitudinal modeling will be carried out for the NFOSI-18 total scores, subscales, a single question within the TSE subscale ("I am bothered by side effects of treatment"), as well as the EQ-5D-5L and EQ-VAS scores for the comparisons among treatment arms using PROC MIXED. Outcomes are PRO post-baseline scores and the predictors are the corresponding baseline PRO score, treatment, randomization stratum, time (treated as a continuous variable), and treatment-by-time interaction. Intercept and time are considered as random effects particular to each patient. All parameter estimates should be obtained using restricted maximum likelihood. The unstructured covariance structure should be used to define covariance between random effects (using option "Type=UN" as a part of the RANDOM statement in PROC MIXED). For the degrees-of-freedom calculations the Kenward and Roger algorithm should be used (using option "ddfm = kr" as a part of the MODEL statement in PROC MIXED). The main analysis will be applied using scheduled assessments from baseline up to EOT (not including EOT), and the sensitivity analysis will be based on all scheduled assessments from baseline including EOT and until 3 years from patient's enrollment.

CCI [redacted]

6.4. Subset Analyses

Subset analyses will be performed for the DDR+ population and for patients unselected for DDR status based on the FAS.

Subset analyses will be performed for PFS per BICR assessment and OS for the subgroups defined below.

The following subgroups will be defined and used for analyses:

- Randomization stratification factor as per IRTs
  - Germline BRCA 1/2 mutational status (gBRCA+ vs gBRCA-)
  - Resection status (adjuvant with >1 mm and ≤1 cm residual disease vs adjuvant with >1 cm residual disease vs neoadjuvant).
- BRCA 1/2 mutational status as determined by Foundation Medicine
  - Positive (Reference)
  - Negative
  - Unknown
- DDR status
  - DDR+
  - DDR-
- Age
  - Age < 65 years (Reference)
  - Age ≥ 65 years
- Race
  - Caucasian / White (Reference)
  - Asian
  - Black/African American
  - Other
- Ethnicity
  - Hispanic/Latino
  - Non-Hispanic/Latino (Reference)
- Pooled Geographical Region
  - North America
  - Europe (Reference)
  - Asia
  - Rest of the World (Australasia, Latin America, Africa and/or Middle East will be included as additional subgroups if including > 10% of the overall randomized population)
- Stage

  - III (Reference)
  - IV
- ECOG Performance Status on or prior to randomization date
  - 0 (Reference)
  - ≥ 1
- CA-125 at baseline
  - ≤2×ULN (Reference)
  - >2×ULN
- Tumor mutation burden
  - Low (reference)
  - High

  The cut-off will be pre-specified using external data before the analysis.
- PD-L1 status at baseline
  - Positive (Reference)
  - Negative

  The cut-off will be pre-specified using external data before the analysis.

Subset analyses for PFS and OS will use the primary censoring rules described in Sections 6.1.1.1 and 6.2.2.2. All the subgroup analyses are exploratory. Treatment arms will be compared for PFS and OS using a 2-sided unstratified log-rank test for each subgroup level and the unstratified HR and its corresponding 95% CI will be computed per subgroup level.

All the subgroup analyses will be exploratory; no adjustment for multiplicity will be performed. In the case of a low number of patients within a category (<5% of the randomized population), the categories will be pooled.

To assess the heterogeneity of treatment effects for PFS and OS across the subgroup levels, two Cox regression model will be fitted with PFS or OS, respectively, as the dependent variable and subgroup, treatment, and with and without the treatment-by-subgroup interaction as explanatory variables.

- Model 1: factors + treatment + subgroup
- Model 2: factors + treatment + subgroup + treatment×subgroup-variable

A p-value for the interaction test (Likelihood Ratio test) will be provided together with the HR and corresponding 95% CI for the interaction model parameter.

The HR for PFS and OS and corresponding 95% CIs for all subgroups will also be presented in a forest plot.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline summaries

The following analyses will be performed for the DDR+ population and for patients unselected for DDR status and will be based on the FAS overall and separately by treatment arm.

6.5.1.1. Demographic characteristics

Demographic characteristics and physical measurements will be summarized by treatment arm using the following information from the 'Screening/Baseline Visit' eCRF pages.

- Demographic characteristics
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown
  - Ethnic origin: Hispanic/Latino (Yes/No)
  - Age (years): summary statistics
  - Age categories:
    - < 65 years, ≥ 65 years
    - < 65, 65-<75, 75-<85, ≥ 85 years
  - Pooled Geographical Region (as applicable):
    - North America
    - Europe
    - Asia
    - Rest of the World (Australasia, Latin America, Africa and/or Middle East will be included as additional pooled geographical regions if including > 10% of the overall randomized population)
  - Geographic Region (as applicable):
    - North America
    - Latin America
    - Western Europe
    - Eastern Europe
    - Middle East
    - Australasia
    - Asia

    - Africa
  - Eastern Cooperative Oncology Group (ECOG) Performance Status: 0, 1, 2, 3, and 4
- Physical measurements
  - Height (cm)
  - Weight (kg)
  - Body Mass Index (BMI) (kg/m^2)
  - Body Surface Area (BSA) (m^2)

Center codes will be used for the determination of the patient's geographic region.

The listing of demographics and baseline characteristics will include the following information: patient identifier, treatment arm, age, sex, race, ethnicity, height (cm), weight (kg), BMI (kg/m^2), BSA (m^2), and ECOG performance status.

6.5.1.2. Medical history

Medical history will be coded using the most current available version of Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized from the 'Significant Medical History' eCRF page. Medical history will be summarized as the numbers and percentages of patients by MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) as summary category. Each patient will be counted only once within each PT or SOC.

Medical history will be displayed in terms of frequency tables: ordered by primary SOC and PT in alphabetical order.

6.5.1.3. Disease characteristics

Information on disease characteristics collected on 'Primary Diagnosis', 'Substance Use' and RECIST eCRF pages will be summarized overall and by treatment arm. Summary statistics will be presented for the following.

From the 'Primary Diagnosis' eCRF page:

- Site of primary tumor
- Primary diagnosis (summarize all categories collected in the 'Primary Diagnosis' eCRF page)
- Time since initial diagnosis to date of randomization (months), defined as (date of randomization – date of initial diagnosis)/30.4375
- Time since histopathological diagnosis (months), defined as (date of randomization – date of histopathological diagnosis)/30.4375

Separately from the RECIST eCRF page based on Investigator assessment and based on BICR assessment:

- Measurable disease (lesions) at baseline (Yes, No, No disease)
- Involved tumor sites at baseline

From the 'Substance Use' eCRF page:

- Smoking history: Never smoker vs current vs former smoker

Listing of disease history will be provided with all relevant data (as collected on the 'Primary Diagnosis' and 'Substance Use' eCRF pages) and derived variables as above.

6.5.1.4. Prior anti-cancer surgery

The prior anti-cancer surgeries are collected under the 'Prior Anti-Cancer Surgery' eCRF page.

The number and percentage of patients with at least one prior anti-cancer surgery will be tabulated. Treatment intent and surgery outcome will also be summarized.

Prior anti-cancer surgeries will be included in the listing of anti-cancer surgeries with a flag to identify prior surgeries. These will include the patient identification number, and all the relevant collected data-fields on the eCRF page.

6.5.2. Study conduct and patient disposition

The following analyses will be performed based on the FAS overall and separately by treatment arm.

6.5.2.1. Patient disposition

The percentages below will be calculated based on the number of patients in the FAS.

- Total number of patients screened overall
- Number of patients who discontinued from the study prior to randomization overall and by the main reason for discontinuation
- Number and percentage of randomized patients in each of the analysis sets defined in Section 4
- Number and percentage of randomized patients with study drug ongoing (separately for each study drug when administered in combination)
- Number and percentage of randomized patients who discontinued study drug overall and by the main reason for discontinuation of study drug (separately for each study drug when administered in combination)
- Number and percentage of patients who entered follow-up

- Number and percentage of patients who discontinued follow-up overall and by the main reason for discontinuation

In addition, a summary of patients who have discontinued all study drugs will be provided.

The results of the randomization algorithm (according to IRT) will be summarized as follows:

- Number and percentage of randomized patients overall, by region (Europe, EEA (required by EudraCT), North America, Latin America, Middle East, Asia, Australasia, Africa), by country within region
- Number and percentage of randomized patients by center
- Number and percentage of randomized patients by randomization strata (IRT)
- Number and percentage of randomized patients by randomization strata (eCRF)
- Cross tabulation: stratum by IRT vs stratum by eCRF
- Cross tabulation: patients randomized (Arm A/Arm B/Arm C) vs patients treated (Arm A/Arm B/Arm C/none)

6.5.2.2. Protocol deviations

All protocol violations that impact the safety of the patients and/or the conduct of a study and/or its evaluation will be reported. These include:

- Patients who are dosed on the study despite not satisfying the inclusion criteria
- Patients who develop withdrawal criteria whilst on the study but are not withdrawn
- Patients who receive the wrong treatment or an incorrect dose
- Patients who receive an excluded concomitant medication
- Deviations from GCP.

The identification of these and other CSR-reportable deviations will be based on the inclusion/exclusion criteria or other criteria presented in the protocol.

6.5.3. Study treatment compliance and exposure

The following analyses will be based on the safety analysis set by treatment arm.

All dosing calculations and summaries will be based on ‘Parenteral – Drug avelumab’ ‘Parenteral – Drug paclitaxel’, ‘Parenteral – Drug carboplatin’, ‘Parenteral – Drug bevacizumab’, and ‘Oral – talazoparib’ eCRFs pages. A listing of study drug administration will be created with the information collected on these eCRF pages.

Cycle definitions for study drugs that are administered in combination apply to all the study drugs in the combination. Ie, cycle is patient-dependent, rather than study-drug-dependent when study drugs are administered in combination.

For Cycle X, actual cycle start date for each patient is

- the earliest start date of dosing in the Cycle X day 1 visit eCRF exposure page, if the patient received study treatment on that visit (ie, any study drug with dose>0 at that visit)
- the first day of assessments in the Cycle X day 1 visit, if the patient did not receive study treatment on that visit (ie, all study drugs had dose=0 at that visit). Use start date in the exposure page if available; if start date is not available then use date of collection of vital signs on Cycle X day 1 visit.

Actual cycle end date for each patient is,

- for all cycles X except the last cycle, actual cycle end date = actual cycle (X+1) start date – 1 day;
- for the last cycle,
    - actual cycle end date = actual cycle start date + 21 days – 1 day if the last cycle occurred in the chemotherapy period;
    - actual cycle end date = actual cycle start date + 42 days – 1 day if the last cycle occurred in the maintenance period;

Cycle duration (weeks) = (actual cycle end date – actual cycle start date + 1)/7

When summarizing exposure for each study drug, only cycles from first dose of study treatment until the last cycle with non-zero dose of at least one of the study drugs should be included.

Exposure may be summarized (per cycle and overall within a study period) as dose received (cumulative dose, actual dose intensity) and as dose received relative to intended dose (relative dose intensity [RDI]).

The information that will be summarized depends on how the study drug is dosed (eg, infusion cyclical, oral daily). The formulae below should be applied to each study drug separately even when study drugs are administered in combination.

The derivations below are provided for the following.

During the chemotherapy period

- Avelumab administered as a 1-hour intravenous (IV) infusion at a dose of 800 mg on Day 1 of each 3-week cycle for 6 cycles

- Paclitaxel administered as 175 mg/m^2 IV over 3 hours on Day 1 of each 3-week cycle for 6 cycles
- Carboplatin dose AUC 5 or AUC 6 IV over 1 hour on Day 1 of each 3-week cycle for 6 cycles
- Bevacizumab administered as an IV infusion at a dose of 15 mg/kg on Day 1 of each 3-week cycle beginning with cycle 2 for adjuvant patients (may begin with cycle 1 if surgery completed >4 weeks prior to randomization), and for neoadjuvant patients, bevacizumab will be given on Day 1 of each 3-week cycle in cycles 1, 2, 5 and 6 for 6 cycles.

During the maintenance period

- Avelumab administered as a 1-hour IV infusion at a dose of 800 mg on Days 1, 15 and 29 of each 6-week cycle for a maximum of 24 months.
- Talazoparib 1 mg administered orally once a day, every day of each 6-week cycle for a maximum of 24 months
- Bevacizumab administered as an IV infusion at a dose of 15 mg/kg on Days 1 and 22 of each 6-week cycle. The maximum duration of treatment is 21 doses total, including the period of initial chemotherapy.

Maintenance period should begin within 4 weeks of the last dose of platinum. The first visit of maintenance therapy will be designated maintenance Cycle 1 Day 1 even if the last chemotherapy cycle was not completed.

Analysis of exposure will be based on the calculated actual dose levels:

- Avelumab - total dose
- Paclitaxel – total dose/m^2
- Carboplatin – AUC
- Bevacizumab – total dose/weight
- Talazoparib – total dose.

6.5.3.1. Exposure to avelumab

The dose level for avelumab is calculated as actual dose administered (mg).

Chemotherapy period

Intended duration of treatment with avelumab (weeks) =

(end date–date of first dose of study drug +1)/7,

where end date = start date of last cycle in chemotherapy period with non-zero dose of study drug + 21 – 1

Duration of exposure to avelumab (weeks) =

(last dose date of avelumab in chemotherapy period – first dose date of avelumab + 21)/7

Cumulative dose in a cycle or overall is the sum of the actual doses of avelumab received in a cycle or overall (in the chemotherapy period), respectively.

Actual Dose Intensity (DI)

- By cycle actual DI (mg/3-week cycle) = [cumulative dose in the cycle (mg)]/[cycle duration (weeks)/3]
- Overall actual DI (mg/3-week cycle) = [overall cumulative dose (mg)] / [intended duration of treatment with avelumab (weeks)/3].

Relative Dose Intensity (RDI)

- Intended DI (mg/ 3-week cycle) = [intended cumulative dose per cycle] / [intended number of 3-weeks in a cycle] = [800 (mg)] / [1 (3-week cycle)] = 800 (mg /3-week cycle)
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [800 (mg/3-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [800 (mg/3-week cycle)]

Maintenance period

Intended duration of treatment with avelumab (weeks) =

(end date–date of first dose of study drug in maintenance period +1)/7,

where end date = start date of last cycle in maintenance period with non-zero dose of study drug + 42 – 1

Duration of exposure to avelumab (weeks) =

(last dose date of avelumab in maintenance period – first dose date of avelumab in maintenance period + 14)/7

Cumulative dose in a cycle or overall is the sum of the actual doses of avelumab received in a cycle or overall (in the maintenance period), respectively

Actual Dose Intensity (DI)

- By cycle actual DI (mg /6-week cycle) = [cumulative dose in the cycle (mg)]/[cycle duration (weeks)/6]
- Overall actual DI (mg/ 6-week cycle) = [overall cumulative dose (mg)] / [intended duration of treatment with avelumab (weeks)/6].

Relative Dose Intensity (RDI)

- Intended DI (mg/ 6-week cycle) = 2400 (mg/ 6-week cycle)
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [2400 (mg/ 6-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [2400 (mg/ 6-week cycle)]

6.5.3.2. Exposure to paclitaxel

The dose level for paclitaxel is calculated as actual dose administered/ m^2 (mg/m^2). The last available weight of the patient on or prior to the day of dosing will be used.

Intended duration of treatment with paclitaxel (weeks) =

(end date–date of first dose of study drug +1)/7,

where end date = start date of last cycle in chemotherapy period with non-zero dose of study drug + 21 – 1

Duration of exposure to paclitaxel (weeks) =

(last dose date of paclitaxel – first dose date of paclitaxel + 21)/7

Cumulative dose in a cycle or overall is the sum of the actual doses of paclitaxel received in a cycle or overall (in chemotherapy period), respectively.

Actual Dose Intensity (DI)

- By cycle actual DI (mg/m^2/3-week cycle) = [cumulative dose in the cycle (mg/m^2)]/[cycle duration (weeks)/3]
- Overall actual DI (mg/m^2/3-week cycle) = [overall cumulative dose (mg/m^2)] / [intended duration of treatment with paclitaxel (weeks)/3].

Relative Dose Intensity (RDI)

- Intended DI (mg/m^2/3-week cycle) = [intended cumulative dose per cycle] / [intended number of 3-weeks in a cycle] = [175 (mg/m^2)] / [1 (3-week cycle)] = 175 (mg/m^2/3-week cycle)
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [175 (mg/m^2/3-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [175 (mg/m^2/3-week cycle)]

6.5.3.3. Exposure to carboplatin

The dose level for carboplatin is calculated as actual AUC administered. The carboplatin dose will be calculated based on the Calvert formula using the actual dose (mg) administered.

Intended duration of treatment with carboplatin (weeks) =

(end date–date of first dose of study drug +1)/7,

where end date = start date of last cycle in the chemotherapy period with non-zero dose of study drug + 21 – 1

Duration of exposure to carboplatin (weeks) =

(last dose date of carboplatin– first dose date of carboplatin+ 21)/7

Cumulative dose in a cycle or overall is the sum of the actual doses of carboplatin received in a cycle or overall (in chemotherapy period), respectively.

Actual Dose Intensity (DI)

- By cycle actual DI (AUC/3-week cycle) = [cumulative dose in the cycle (AUC)]/[cycle duration (weeks)/3]
- Overall actual DI (AUC /3-week cycle) = [overall cumulative dose (AUC)] / [intended duration of treatment with carboplatin (weeks)/3].

Relative Dose Intensity (RDI)

- Intended DI (AUC/3-week cycle) = [intended cumulative dose per cycle] / [intended number of 3-weeks in a cycle] = [d (AUC)] / [1 (3-week cycle)] = d (AUC/3-week cycle)
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [d (AUC/3-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [d AUC /3-week cycle)]

  where d= 5 or 6 depending on the glomerular filtration rate (GFR) method used for the Cycle 1 Day 1 carboplatin dose.

6.5.3.4. Exposure to bevacizumab

The dose level for bevacizumab is calculated as actual dose administered/weight (mg/kg). The last available weight of the patient on or prior to the day of dosing will be used.

Chemotherapy period

Intended duration of treatment with bevacizumab (weeks) =

(end date–date of first dose of bevacizumab +1)/7,

where end date = start date of last cycle in chemotherapy period with non-zero dose of study drug + 21 – 1

Duration of exposure to bevacizumab (weeks) =

(last dose date of bevacizumab in the chemotherapy period – first dose date of bevacizumab + 21)/7

Cumulative dose in a cycle or overall is the sum of the actual doses of bevacizumab received in a cycle or overall (in chemotherapy period), respectively.

Actual Dose Intensity (DI)

- By cycle actual DI (mg/kg/3-week cycle) = [cumulative dose in the cycle (mg/kg)]/[cycle duration (weeks)/3]
- Overall actual DI (mg/kg/3-week cycle) = [overall cumulative dose (mg/kg)] / [intended duration of treatment with bevacizumab (weeks)/3].

Relative Dose Intensity (RDI)

- Intended DI (mg/kg/3-week cycle) = [intended cumulative dose per cycle] / [intended number of 3-weeks in a cycle] = [d (mg/kg / 3-week cycle]

  where d=15 for adjuvant patients and d=10 for neoadjuvant patients.
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [d (mg/kg/3-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [d (mg/kg/3-week cycle)]

Maintenance period

Intended duration of treatment with bevacizumab (weeks) =

(end date–date of first dose of bevacizumab in maintenance period +1)/7,

where end date = start date of last cycle in maintenance period with non-zero dose of study drug + 42 – 1

Duration of exposure to bevacizumab (weeks) =

(last dose date of bevacizumab in maintenance period – first dose date of bevacizumab in maintenance period + 21)/7

Cumulative dose in a cycle or overall is the sum of the actual doses of bevacizumab received in a cycle or overall (in maintenance period), respectively.

Actual Dose Intensity (DI)

- By cycle actual DI (mg/kg/6-week cycle) = [cumulative dose in the cycle (mg/kg)]/[cycle duration (weeks)/6]
- Overall actual DI (mg/kg/6-week cycle) = [overall cumulative dose (mg/kg)] / [intended duration of treatment with Drug D (weeks)/6].

Relative Dose Intensity (RDI)

- Intended DI (mg/kg/6-week cycle) = [intended cumulative dose per cycle] / [intended number of 6-weeks in a cycle] = [2 x 15 (mg/kg)] / [1 (6-week cycle)] = 30 (mg/kg/6-week cycle)
- By cycle RDI (%) = 100 × [by cycle actual DI] / [intended DI]
  = 100 × [by cycle actual DI] / [30 (mg/kg/6-week cycle)]
- Overall RDI (%) = 100 × [overall actual DI] / [intended DI]
  = 100 × [overall actual DI] / [30 (mg/kg/6-week cycle)]

6.5.3.5. Exposure to talazoparib

The dose level is calculated as actual dose administered (mg/day).

Intended duration of treatment with talazoparib (weeks) = (end date – date of first dose of talazoparib in maintenance period +1)/7,

where end date = date of last dose of talazoparib.

Duration of exposure to talazoparib (weeks) =

(last dose date of talazoparib – first dose date of talazoparib + 1)/7

Cumulative dose is the sum of the actual doses of talazoparib received in the study.

Actual Dose Intensity (DI)

- Overall actual DI (mg/week) = [overall cumulative dose (mg)] / [intended treatment duration (weeks)]

Relative Dose Intensity (RDI)

- RDI (%) = 100 × [overall cumulative dose] / [intended cumulative dose per week × number of weeks from first dose of talazoparib to last dose of talazoparib]

  = 100 × [overall cumulative dose] / [7 × d × duration of exposure to talazoparib in weeks]

  where d is the dose for talazoparib.

6.5.3.6. Dose reductions

Applicable to avelumab, paclitaxel, carboplatin, and bevacizumab. Dose reduction is defined as actual non-zero dose < 90% of the planned dose.

Applicable to talazoparib. Dose reduction is defined as a change to a non-zero dose level lower than that planned in the protocol.

The number and percentage of patients with at least one dose reduction as well as a breakdown of the number of dose reductions (1, 2, 3, ≥4) will be summarized.

6.5.3.7. Dose interruptions

Applicable to talazoparib only.

An interruption is defined a 0 mg dose administered on one or more days for talazoparib. What follows defines how dose interruptions will be counted in the case of multiple dose interruptions.

- If an interruption occurs consecutively for at least two days due to the same reason, then it will be counted only once (example: If the actual dose on days 1-3 is 1 mg and actual dose on days 4-5 is 0 mg and dose interruption on days 4-5 is due to AE, then the total number of dose interruptions is 1).
- If an interruption occurs consecutively for at least two days due to different reasons, then it will be counted for each reason (example: If the actual dose on days 1-3 is 1 mg and actual dose on days 4-5 is 0 mg and dose interruption on day 4 is due to AE and dose interruption on day 5 is due to dosing error, then the total number of dose interruptions is 2).
- If an interruption occurs for more than one day due to the same reason, but the days are not consecutive, ie there is at least one dosing day in between, then each dose interruption will be counted as a different occurrence (example: If the actual dose on days 1, 3 and 5, is 1 mg and actual dose on days 2 and 4 is 0 mg, and dose interruptions on day 2 and 4 are both due to dosing error, the total number of dose interruptions is 2).

A dose interruption is not considered a dose reduction.

The number and percentage of patients with dose interruptions and the corresponding reasons will be summarized.

6.5.3.8. Dose delays

Applicable to avelumab, paclitaxel, carboplatin, and bevacizumab.

Dose Delay is the difference between the actual time between two consecutive non-zero doses and the planned time between the same two consecutive non-zero doses.

Dose Delay for Dose x (days) = Date of Dose x – Date of Dose (x-1) – Planned days between two consecutive doses, where planned days are:

- 21 for carboplatin,
- 21 for paclitaxel,
- 21 for bevacizumab in both chemotherapy and maintenance periods,
- 21 for avelumab in the chemotherapy period, 14 for avelumab in the maintenance period.

For patients taking IDS, the interval encompassing the IDS date (ie Dose (x-1) Date < surgery < Dose x Date) will not be considered for dose delay assessment.

Dose delays will be grouped into the following categories:

- No delay
- 1-3 days delay
- 4-6 days delay
- 7 or more days delay

For example, for avelumab administered on a 2-week schedule, if one patient receives avelumab on Day 1, then the next avelumab administration date will be on Day 15; however, if the patient receives avelumab at Day 16, 17, or 18, this is considered as 1-3 days delay.

No delay and 1-3 days delay will also be summarized together.

The number and percentage of patients with delayed study drug administration and maximum length of delay, ie, the worst case of delay if patients have multiple dose delays will be summarized.

6.5.3.9. Infusion rate reductions

Applicable to avelumab, paclitaxel, carboplatin, and bevacizumab.

The number and percentage of patients with at least one infusion rate reduction of ≥50% compared to the first infusion rate reported in the eCRF as well as the frequency of patients with 1, 2, 3, or ≥4 infusion rate reductions of ≥ 50% will be summarized.

6.5.3.10. Infusion interruptions

Applicable to avelumab, paclitaxel, carboplatin, and bevacizumab.

An infusion interruption is defined as an infusion that is stopped and re-started on the same day (ie, for a visit more than one infusion start time and infusion end time are recorded).

The number and percentage of patients with at least one infusion interruption as well as the frequency of patients with 1, 2, 3, or ≥4 infusion interruptions will be summarized.

6.5.4. Concomitant medications and non-drug treatments

The following analyses will be based on the safety analysis set by treatment arm.

Concomitant medications are medications, other than study medications, which started prior to first dose date of study treatment and continued on on-treatment period as well as those started during the on-treatment period. **Prior medications** are medications, other than study medications and pre-medications for study drug, which are started before the first dose of study treatment.

Prior and concomitant medications will be summarized from the ‘General Concomitant Medications’ eCRF page. Pre-medications for study drug will also be summarized separately from the ‘Pre-Medication Treatment’ eCRF page.

Summary of prior medications, summary of concomitant medications and summary of pre-medications will include the number and percentage of patients by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term. A patient will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. If any prior or concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have ATC classification level 2 coded term, it will be summarized under 'Unavailable ATC classification' category.

A listing of prior medications and a listing of concomitant medications will be created with the relevant information collected on the 'General Concomitant Medications' eCRF page. A listing of pre-medications will be created with the relevant information collected on the 'Pre-Medication Treatment' eCRF page.

All concurrent procedures, which were undertaken any time during the on-treatment period, will be listed according to the eCRF page 'General Non-drug Treatments'.

A listing of concurrent procedures will be created with the relevant information collected on the 'General Non-drug Treatments' eCRF page.

6.5.5. Subsequent anti-cancer therapies

The following analyses will be based on the FAS by treatment arm.

Anti-cancer treatment will be provided in a data listing with data retrieved from 'Follow-up Cancer Therapy', 'Concomitant Radiation Therapy', 'Follow-up Radiation Therapy', 'On-Study Cancer Surgery', and 'Follow-up Cancer Surgery' eCRF pages.

- Listing of anti-cancer drug therapies
- Listing of anti-cancer radiotherapy
- Listing of anti-cancer surgeries

Number and percentage of patients with any anti-cancer therapy after discontinuation will be tabulated overall and by type of therapy based on the data collected from the 'Follow-up Cancer Therapy', 'Follow-up Radiation Therapy' and 'Follow-up Cancer Surgery' eCRF pages.

6.6. Safety Summaries and Analyses

The Safety Analysis Set will be the primary population for safety evaluations. Summaries of AEs and other safety parameters will be based on the safety analysis set by treatment arm.

6.6.1. Adverse events

Treatment-emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period for the first time, or if the worsening of an event is during the on-treatment period as defined in Section 3.5.1.

All analyses described will be based on TEAEs (started during the on-treatment period) if not otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not). AEs outside the on-treatment period will be flagged in the listings.

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (ie, no answer to the question 'Relationship with study treatment'). Related AEs are those related to any study drug (ie, at least one of the study drugs).
- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious Adverse Event = Yes).
- **Adverse Events Leading to Dose Reduction:** adverse events leading to dose reduction of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Dose reduced).
- **Adverse Events Leading to Interruption of Study Treatment:** adverse events leading to interruption of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug interrupted). The eCRF does not allow for a clear separation between interruption of an infusion and delays of administration for a parenteral drug as both are recorded using the same term on the eCRF ("Drug interrupted"). IRRs will be excluded in the analysis of AEs leading to Drug Interruption in case they only led to an interruption of the infusion.
- **Adverse Events Leading to Permanent Treatment Discontinuation:** adverse events leading to permanent discontinuation of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug withdrawn).
- **Adverse Events Leading to Death:** adverse event leading to death (as recorded on the AE eCRF page, Outcome = Fatal, as well as AEs of Grade 5).
- **Immune-related Adverse Events (irAE):** irAEs (as identified according to the methodology outlined in Appendix 1 for a pre-specified search list of MedDRA PTs documented in the Safety Review Plan and finalized for analysis of the current studies data prior to DB lock)
- **Infusion-related Reactions (IRR)**: IRRs (as identified according to the methodology outlined in Appendix 2 for a pre-specified search list of MedDRA PTs documented in the Safety Review Plan and finalized for analysis of the current studies data prior to DB lock).
- Complications Related to Surgery (CRS): CRS as collected on dedicated AE page

Unless otherwise specified, AEs will be summarized by number and percentage of patients with the AE in the category of interest as described above, by treatment arm, primary SOC and PT in decreasing frequency based on the frequencies observed for Arm A.

Each patient will be counted only once within each SOC or PT. If a patient experiences more than one AE within a SOC or PT for the same summary period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

6.6.1.1. All adverse events

Adverse events will be summarized by worst severity (according to NCI-CTCAE version 4.03) per patient, using the latest version of MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) body term as Body System category.

In case a patient has events with missing and non-missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

The following tables will be created:

- The overall summary of AEs table will include the frequency (number and percentage) of patients with each of the following by treatment arm:
  - TEAEs
  - TEAEs, Grade ≥ 3
  - Related TEAEs
  - Related TEAEs, Grade ≥ 3
  - TEAEs leading to dose reduction of avelumab
  - TEAEs leading to dose reduction of paclitaxel
  - TEAEs leading to dose reduction of carboplatin
  - TEAEs leading to dose reduction of bevacizumab
  - TEAEs leading to dose reduction of talazoparib
  - TEAEs leading to interruption of avelumab
  - TEAES leading to interruption of paclitaxel
  - TEAEs leading to interruption of carboplatin
  - TEAEs leading to interruption of bevacizumab
  - TEAEs leading to interruption of talazoparib
  - TEAEs leading to discontinuation of avelumab
  - TEAEs leading to discontinuation of paclitaxel
  - TEAEs leading to discontinuation of carboplatin

- TEAEs leading to discontinuation of bevacizumab
- TEAEs leading to discontinuation of talazoparib
- TEAEs leading to discontinuation of any study drug
- TEAEs leading to discontinuation of all study drugs
- Related TEAEs leading to discontinuation of avelumab
- Related TEAEs leading to discontinuation of paclitaxel
- Related TEAEs leading to discontinuation of carboplatin
- Related TEAEs leading to discontinuation of bevacizumab
- Related TEAEs leading to discontinuation of talazoparib
- Related TEAEs leading to discontinuation of any study drug
- Related TEAEs leading to discontinuation of all study drugs
- Serious TEAEs
- Related Serious TEAEs
- TEAEs leading to death
- Related TEAEs leading to death
- irAEs
- IRRs

- TEAEs by SOC and PT and worst grade
- Related TEAEs by SOC and PT and worst grade
- TEAEs leading to death by SOC and PT
- Related TEAEs leading to death by SOC and PT
- TEAEs Excluding SAEs, with frequency ≥ 5% in any treatment arm by SOC and PT

6.6.1.2. Adverse events leading to dose reduction

The frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to dose reduction of each study drug by treatment arm:

- TEAEs leading to dose reduction of avelumab by SOC and PT
- TEAEs leading to dose reduction of paclitaxel by SOC and PT
- TEAEs leading to dose reduction of carboplatin by SOC and PT
- TEAEs leading to dose reduction of bevacizumab by SOC and PT

- TEAEs leading to dose reduction of talazoparib by SOC and PT

The listing of all AEs leading to dose reduction will also be provided with the relevant information.

6.6.1.3. Adverse events leading to interruption of study treatment

The eCRF does not allow for a clear separation between interruption of an infusion and delays of administration for a parenteral drug as both are recorded using the same term on the eCRF ("Drug interrupted"). IRRs will be excluded in the analysis of AEs leading to Drug Interruption in case they only led to an interruption of the infusion (ie, did not lead to a dose reduction or a dose delay).

As such, for parenteral drugs, AEs leading to interruption will be defined as AEs identified in the AE eCRF page with an action taken with study treatment of 'drug interrupted' excluding

- IRRs that occurred on the day of infusion with ≥90% of the planned dose given (ie IRRs that did not lead to a dose reduction) and subsequent administration of study drug had no delay (as defined in Section 6.5.3.8). These IRRs will be considered as IRRs leading to interruption of infusion.
- IRRs occurring on the day after infusion and subsequent dose administration had no delay (as defined in Section 6.5.3.8).

For talazoparib, AEs leading to interruption will be defined as all AEs identified in the AE eCRF page with an action taken with study treatment of 'drug interrupted'.

The frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to interruption of each study drug by treatment arm:

- TEAEs leading to interruption of avelumab by SOC and PT
- TEAEs leading to interruption of paclitaxel by SOC and PT
- TEAEs leading to interruption of carboplatin by SOC and PT
- TEAEs leading to interruption of bevacizumab by SOC and PT
- TEAEs leading to interruption of talazoparib by SOC and PT

The listing of all AEs leading to interruption of study treatment will also be provided with the relevant information.

In addition, the frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to interruption of each study drug by treatment arm:

- TEAEs leading to both interruption and dose reduction of avelumab by SOC and PT
- TEAEs leading to both interruption and dose reduction of paclitaxel by SOC and PT
- TEAEs leading to both interruption and dose reduction of carboplatin by SOC and PT

- TEAEs leading to both interruption and dose reduction of bevacizumab by SOC and PT
- TEAEs leading both interruption and dose reduction of talazoparib by SOC and PT

This summary will take into account PTs with both actions as defined in Section 6.6.1, even though the actions may be captured for different PT records (ie, different onset for the PT with action "drug interrupted" and the PT with action "dose reduced".

6.6.1.4. Adverse events leading to discontinuation of study treatment

The frequency (number and percentage) of patients with each of the following will be presented for TEAEs leading to permanent discontinuation of each study drug and study treatment, by treatment arm:

- TEAEs leading to discontinuation of avelumab by SOC and PT
- Related TEAEs leading to discontinuation of avelumab by SOC and PT
- TEAEs leading to discontinuation of paclitaxel by SOC and PT
- Related TEAEs leading to discontinuation of paclitaxel by SOC and PT
- TEAEs leading to discontinuation of carboplatin by SOC and PT
- Related TEAEs leading to discontinuation of carboplatin by SOC and PT
- TEAEs leading to discontinuation of bevacizumab by SOC and PT
- Related TEAEs leading to discontinuation of bevacizumab by SOC and PT
- TEAEs leading to discontinuation of talazoparib by SOC and PT
- Related TEAEs leading to discontinuation of talazoparib by SOC and PT
- TEAEs leading to discontinuation of any study drug by SOC and PT
- Related TEAEs leading to discontinuation of any study drug by SOC and PT
- TEAEs leading to discontinuation of all study drugs by SOC and PT
- Related TEAEs leading to discontinuation of all study drugs by SOC and PT

The listing of all AEs leading to treatment discontinuation will also be provided with the relevant information.

6.6.2. Deaths

The frequency (number and percentage) of patients in the safety analysis set who died and who died within 30 days after last dose of study treatment as well as the reason for death, will be tabulated based on information from the 'Notice of Death' and 'Survival Follow-Up' eCRFs, by treatment arm.

- All deaths

- Deaths within 30 days after last dose of study treatment
- Reason for Death
  - Disease progression
  - Study treatment toxicity
  - AE not related to study treatment
  - Unknown
  - Other.

In addition, date and cause of death will be provided in individual patient data listing together with selected dosing information (study treatment received, date of first / last administration, dose) and will include the following information:

- AEs with fatal outcome (list preferred terms of AEs with outcome=Fatal, as well as AEs of Grade 5),
- Flag for death within 30 days of last dose of study treatment.

6.6.3. Serious adverse events

The frequency (number and percentage) of patients with each of the following will be presented for treatment-emergent SAEs by treatment arm:

- SAEs by SOC and PT
- Related SAEs by SOC and PT

The listings of all SAEs will also be provided with the relevant information with a flag for SAEs with onset outside of the on-treatment period.

6.6.4. Other significant adverse events

The frequency (number and percentage) of patients with each of the following will be presented for irAEs, by treatment arm:

- irAEs leading to death, by Cluster and PT
- irAEs, by Cluster and PT
- irAEs, Grade ≥ 3, by Cluster and PT
- irAEs leading to discontinuation of avelumab, by Cluster and PT
- irAEs leading to discontinuation of paclitaxel, by Cluster and PT
- irAEs leading to discontinuation of carboplatin, by Cluster and PT
- irAEs leading to discontinuation of bevacizumab, by Cluster and PT

- irAEs leading to discontinuation of talazoparib, by Cluster and PT
- irAEs leading to discontinuation of any study drug, by Cluster and PT
- irAEs leading to discontinuation of all study drugs, by Cluster and PT
- Serious irAEs, by Cluster and PT

The listing of all irAEs will also be provided with the relevant information with a flag for irAEs with onset outside of the on-treatment period.

The frequency (number and percentage) of patients with each of the following will be presented for IRRs, by treatment arm:

- IRRs leading to death, by PT
- IRRs, by PT
- IRRs, Grade ≥ 3, by PT
- IRRs leading to discontinuation of avelumab, by PT
- IRRs leading to discontinuation of paclitaxel, by PT
- IRRs leading to discontinuation of carboplatin, by PT
- IRRs leading to discontinuation of bevacizumab, by PT
- IRRs leading to discontinuation of any study drug, by PT
- IRRs leading to discontinuation of all study drugs, by PT
- Serious IRRs, by PT
- Time related to first onset of an IRR (infusion 1, infusion 2, infusion 3, infusion 4 or later). For IV study drugs administered in combination the infusion numbers are those associated with the regimen, rather than the individual study drugs.

The listing of all IRRs will also be provided with the relevant information with a flag for IRRs with onset outside of the on-treatment period.

The frequency (number and percentage) of patients with each of the following will be presented for treatment-emergent CRSs, by treatment arm:

- CRSs leading to death, by PT
- CRSs, by PT
- CRSs, Grade ≥ 3, by PT
- CRSs leading to discontinuation of avelumab, by PT

- CRSs leading to discontinuation of paclitaxel, by PT
- CRSs leading to discontinuation of carboplatin, by PT
- CRSs leading to discontinuation of bevacizumab, by PT
- CRSs leading to discontinuation of talazoparib, by PT
- CRSs leading to discontinuation of any study drug, by PT
- CRSs leading to discontinuation of all study drugs, by PT
- Serious CRSs, by PT

The listing of all CRSs will also be provided with the relevant information with a flag for CRSs with onset outside of the on-treatment period.

In addition, the following analyses will be presented for Tier-1 (irAEs and IRRs as defined in Appendices 1 and 2) and Tier-2 events separately. P-values and CIs for risk difference will be calculated based on the unconditional exact method by Santner and Snell (1980). No additional analyses will be presented for Tier-3 AEs.

- Frequency (number and percentage) of patients with each of the following by treatment arm and PT or Clustered Term:
  - Tier-2 AEs
  - Tier-2 AEs Grade ≥ 3
- Point estimate for risk difference and 95% CI for risk difference (experimental arm vs comparator arm) for each of the following by PT or Clustered Term:
  - irAEs
  - irAEs Grade ≥ 3
  - IRRs
  - IRRs Grade ≥ 3
  - Tier-2 AEs
  - Tier-2 AEs Grade ≥ 3
- 2-sided p-value associated with risk difference (experimental arm vs comparator arm) for each of the following by PT or Clustered Term:
  - irAEs
  - irAEs Grade ≥ 3
  - IRRs

- IRRs Grade ≥ 3

- The p-values and CIs reported are not adjusted for multiplicity and should be used for screening purposes only. The 95% CIs are provided to help gauge the precision of the estimates for the risk difference and should be used for estimation purposes only.

6.6.5. Laboratory data

6.6.5.1. Hematology and chemistry parameters

Laboratory results will be classified according to the NCI-CTCAE criteria version 4.03. Non-numerical qualifiers (with the exception of fasting flags) will not be taken into consideration in the derivation of CTCAE criteria (eg, hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived). Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: below normal limit, within normal limits, and above normal limit (according to the laboratory normal ranges).

Quantitative data will be summarized using simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each nominal visit over time (unscheduled measurements would therefore not be included in these summaries as described in Section 5.2.9). End of Treatment visit laboratory results will be summarized separately. The changes computed will be the differences from baseline. Qualitative data based on reference ranges will be described according to the categories (ie, Low, Normal, High).

Abnormalities classified according to NCI-CTCAE toxicity grading version 4.03 will be described using the worst grade. For those parameters which are graded with two toxicities such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (eg, hypokalemia) grades at baseline and post baseline will be set to 0 when the variables are derived for summarizing high direction toxicity (eg, hyperkalemia), and vice versa.

For **WBC differential counts** (total neutrophil [including bands], lymphocyte, monocyte, eosinophil, and basophil counts), the absolute value will be used when reported. When only percentages are available (this is mainly important for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

Derived differential absolute count = (WBC count) × (Differential %value / 100)

If the range for the differential absolute count is not available (only range for value in % is available) then Grade 1 will be attributed to as follows:

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 800/mm3

- Neutrophil count decreased
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 1500/mm3

For **calcium**, CTCAE grading is based on Corrected Calcium and Ionized Calcium (CALCIO). Corrected Calcium is calculated from Albumin and Calcium as follows

Corrected calcium (mmol/L) = measured total Calcium (mmol/L) + 0.02 (40 - serum albumin [g/L])

Liver function tests: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total bilirubin (TBILI) are used to assess possible drug induced liver toxicity. The ratios of test result over upper limit of normal (ULN) will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of patients with each of the following during the on-treatment period will be summarized by treatment arm:

- ALT ≥ 3×ULN, ALT ≥ 5×ULN, ALT ≥ 10×ULN, ALT ≥ 20×ULN
- AST ≥ 3×ULN, AST ≥ 5×ULN, AST ≥ 10×ULN, AST ≥ 20×ULN
- (ALT or AST) ≥ 3×ULN, (ALT or AST) ≥ 5×ULN, (ALT or AST) ≥ 10×ULN, (ALT or AST) ≥ 20×ULN
- TBILI ≥ 2×ULN
- Concurrent ALT ≥ 3×ULN and TBILI ≥ 2×ULN
- Concurrent AST ≥ 3×ULN and TBILI ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and ALP > 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and TBILI ≥ 2×ULN and (ALP ≤ 2×ULN or missing)

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, ie, a patient with an elevation of AST ≥10×ULN will also appear in the categories ≥5×ULN and ≥3×ULN. Liver function elevation and possible Hy's Law cases will be summarized using frequency counts and percentages.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created, with different symbols for different treatment arms, by graphically displaying

- peak serum ALT(/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT=3×ULN and total bilirubin=2×ULN.
- peak serum AST(/ULN) vs peak total bilirubin (/ULN) including reference lines at AST=3×ULN and total bilirubin=2×ULN.

In addition, a listing of all TBILI, ALT, AST and ALP values for patients with a post-baseline TBILI ≥ 2×ULN, ALT≥ 3×ULN or AST ≥ 3×ULN will be provided.

Parameters with NCI-CTC grades available:

The laboratory toxicities will be tabulated using descriptive statistics (number of patients and percentages) during the on-treatment period. The denominator to calculate percentages for each laboratory parameter is the number of patients evaluable for CTCAE grading (ie those patients for whom a Grade 0, 1, 2, 3, or 4 can be derived).

- The summary of laboratory parameters by CTCAE grade table will include number and percentage of patients with Grade 1, 2, 3, 4, Grade 3/4, and any grade (Grades 1-4), laboratory abnormalities during the on-treatment period.
- The shift table will summarize baseline CTCAE grade versus the worst on-treatment CTCAE grade. The highest CTCAE grade during the on-treatment period is considered as the worst grade for the summary.

The above analyses apply to hematology and chemistry evaluations which can be graded per CTCAE, ie:

- Hematology:

  Hemoglobin (HB), Leukocytes (white blood cell decreased), Lymphocytes (lymphocyte count increased/decreased), Neutrophils / Absolute Neutrophils Count (ANC) (neutrophil count decreased), Platelet Count (PLT) (platelet count decreased).

- Serum Chemistry:

  Albumin (hypoalbuminemia), Alkaline Phosphatase (alkaline phosphatase increased), Alanine Aminotransferase (ALT) (ALT increased), Amylase (serum amylase increased), Aspartate Aminotransferase (AST) (AST increased), Total Bilirubin (blood bilirubin increased, Creatinine (creatinine increased), Creatine Kinase (CPK increased), Potassium (hypokalemia/ hyperkalemia), Sodium (hyponatremia/ hypernatremia), Magnesium (hypomagnesemia/hypermagnesemia), Calcium (hypocalcemia/ hypercalcemia), Glucose (hypoglycemia/hyperglycemia), Gamma Glutamyl Transferase (GGT) (GGT increased), Lipase (lipase increased), Phosphates (hypophosphatemia).

Parameters with NCI-CTC grades not available:

Hematology and chemistry evaluations which cannot be graded per CTCAE criteria will be summarized as frequency (number and percentage) of patients with:

- shifts from baseline normal to at least one result above normal during on-treatment period

- shifts from baseline normal to at least one result below normal during on-treatment period

In this study, these apply to the following parameters:

- Serum Chemistry: Chloride, Uric Acid,

6.6.5.2. Other laboratory parameters

All other parameters collected on the eCRF will be listed in dedicated listings presenting all corresponding collected information on the eCRF.

- Coagulation: activated partial thromboplastin time (aPTT), prothrombin time (PT) or international normalized ratio (INR).
- Thyroid function: TSH, free T4
- Urinalysis:-Protein, glucose, blood
- Other parameters: ACTH, HBV, Anti-HCV antibody. If Anti-HCV antibody positive, then HCV RNA testing
- Pregnancy test

The listings of laboratory results will be provided for all laboratory parameters. The listings will be sorted by parameters and assessment dates or visits for each patient. Laboratory values that are outside the normal range will also be flagged in the data listings, along with corresponding normal ranges. A listing of CTCAE grading will also be generated for those laboratory tests.

In addition, listings of abnormal values will be provided for hematology, chemistry, urinalysis, coagulation parameters. If there is at least one abnormal assessment for any parameter, all the data for that laboratory parameter will be included into the listing.

For all tests not mentioned above but present in the clinical data, a listing of patients with at least one result for the relevant test will be provided.

6.6.6. Vital signs

Weight for the purposes of dose calculation will be recorded at screening and within 3 days pre-dose Day 1 of each cycle. Weight will not be collected at End of Treatment. Height will be measured at screening only.

Vital sign summaries will include all vital sign assessments from the on-treatment period. All vital sign assessments will be listed, and those collected outside the on-treatment period will be flagged in the listing.

All vital sign parameters will be summarized using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each visit over time. End of Treatment visit will be summarized separately. The changes computed will be the differences from baseline.

6.6.7. Electrocardiogram

All patients require a single ECG measurement at screening (clinically significant abnormal findings in baseline ECGs will be recorded as medical history). Additional ECGs will be performed as clinically indicated. Clinically significant findings seen on follow-up ECGs should be recorded as adverse events. All ECG assessments will be listed.

6.6.8. Physical examination

Patient physical assessments, scheduled and performed, will be listed.

6.6.9. ECOG performance status

The ECOG shift from baseline to highest score during the on-treatment period will be summarized by treatment arm. ECOG performance status with shift from ECOG=0 or 1 to ECOG 2 or higher will also be presented in a data listing.

7. INTERIM ANALYSES

7.1. Introduction

The goals of the interim analyses are to allow early stopping of treatment arm(s) for futility or efficacy. The interim analysis of PFS and OS will be performed as described in Sections 5.1.1 and 5.1.2 using the methodology described in Sections 6.1.1.1 and 6.2.2.1 for PFS and Section 6.2.2.2 for OS.

At the time of the interim analysis for PFS in the DDR+ population, hierarchical testing of other secondary endpoints will also be carried out as outlined in Section 5.1.1, with the analyses being performed by an independent statistician. Unblinded results from the interim analysis will not be communicated to the Sponsor's clinical team or to any party involved in the study conduct (apart from the independent statistician and E-DMC members) until the E-DMC has determined that either (i) PFS analysis in the DDR+ population has crossed the pre-specified boundary for efficacy or (ii) the study needs to be terminated due to any cause, including futility or safety reasons. Further details will be described in the E-DMC charter.

At the time of final analysis of PFS in the DDR+ population, the analyses of other secondary endpoints will be performed by the Sponsor's clinical team. All patients will continue to be followed for OS until the final OS analysis (or earlier if OS reaches statistical significance at an interim analysis).

7.2. Interim Analyses and Summaries

At each analysis time point, the critical boundaries for the group sequential test will be derived from the predefined spending function(s) as described in Section 5.1. The calculations of boundaries will be performed using EAST.

7.2.1. Interim analysis for PFS in the DDR+ population

Throughout this section, references to PFS pertain to PFS based on BICR assessment in the DDR+ population.

Let $u(t_1)$ and $u(t_F)$ denote the upper critical boundaries based on the test statistics Z_1 and Z_F for efficacy at the interim and the final analysis, respectively, and let $l(t_1)$ and $l(t_F)$ denote the lower critical boundary for futility at the interim and final analysis, respectively. For the final analysis, $l(t_F)=u(t_F)$.

The critical values $u(t_1)$ and $l(t_1)$ for the interim analysis of PFS are determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1) \quad \text{and} \quad P_a(Z_1 \leq l(t_1)) = \beta(t_1) \ ,$$

where P_0 and P_a denote the probabilities under the null hypothesis and the alternative hypothesis, respectively, and $\alpha(t_1)$ and $\beta(t_1)$ denote the α and β spent, respectively, based on the predefined spending functions at information fraction t_1 (t_1 is calculated as the ratio of the number of PFS events observed at the time of the cut-off for the interim analysis and the total number of PFS events targeted for the final analysis).

The boundary for the final efficacy analysis will be calculated such that

$$\alpha(t_1) + P_0(Z_1 < u(t_1), Z_F \geq u_F) = 0.025$$

As described in Section 5.1.2, if the number of PFS events in the final analysis deviates from the target number of PFS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analysis and the actual correlation between the two test statistics Z_1 and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.025.

7.2.2. Interim analysis for PFS in patients unselected for DDR status

Throughout this section, references to PFS pertain to PFS based on BICR assessment in patients unselected for DDR status.

The interim analysis for PFS for patients unselected for DDR status will be performed if H_{01} is rejected. No futility analysis will be performed for PFS in patients unselected for DDR status.

Let $u(t_1)$ and $u(t_F)$ denote the upper critical boundaries based on the test statistics Z_1 and Z_F for efficacy at the interim and the final analysis, respectively.

The critical value $u(t_1)$ for the interim analysis of PFS are determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1) \ ,$$

where P_0 denotes the probabilities under the null hypothesis, and $\alpha(t_1)$ denotes the α spent, based on the predefined spending functions at information fraction t_1 (t_1 is calculated as the ratio of the number of PFS events observed at the time of the cut-off for the interim analysis and the total number of PFS events targeted for the final analysis).

The boundary for the final efficacy analysis will be calculated such that

$$\alpha(t_1) + P_0(Z_1 < u(t_1), Z_F \geq u_F) = 0.003$$

As described in Section 5.1.2, if the number of PFS events in the final analysis deviates from the target number of PFS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analysis and the actual correlation between the two test statistics Z_1 and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.003.

7.2.3. Interim analysis for OS in the DDR+ population

Throughout this section, references to OS pertain to OS in the DDR+ population.

The interim analysis for OS in the DDR+ population will be performed if H_{01} is rejected. No futility analysis will be performed for OS in the DDR+ population.

Let $u(t_i)$ and $u(t_F)$ denote the upper critical boundaries based on the test statistics Z_i and Z_F for efficacy at the i^{th} interim and the final analysis, respectively, where i=1, 2, 3.

In what follows P_0 denotes the probability under the null hypothesis, and $\alpha(t_i)$ denotes the α spent based on the predefined α-spending function at information fraction t_i (t_i is calculated as the ratio of the number of OS events observed at the time of the cut-off for the i^{th} interim analysis and the total number of OS events targeted for the final analysis.

For each comparison, the critical value $u(t_1)$ for the 1^{st} interim analysis of OS is determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1).$$

Critical boundaries for the additional interim analyses and the final analysis of OS are calculated recursively as follows for each comparison

u(t2) is derived such that $\alpha(t_1) + P_0(Z_1 < u(t_1), Z_2 \geq u(t_2)) = \alpha(t_2)$,

u(t3) is derived such that $\alpha(t_2) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 \geq u(t_3)) = \alpha(t_3)$

The boundary for the final efficacy analysis is derived such that

$$\alpha(t_3) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 < u(t_3), Z_F \geq u_F) = 0.022$$

As described in Section 5.1.2, if the number of OS events in the final analysis deviates from the target number of OS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analyses and the actual correlation between the four test statistics Z_i (i=1, 2, 3) and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.022.

7.2.4. Interim analysis for OS in patients unselected for DDR status

Throughout this section, references to OS pertain to OS in patients unselected for DDR status.

The interim analysis for OS for patients unselected for DDR status will be performed according to the testing strategy described in Section 5.1.1. No futility analysis will be performed for OS in patients unselected for DDR status.

In what follows P_0 denotes the probability under the null hypothesis, and $\alpha(t_i)$ denotes the α spent based on the predefined α-spending function at information fraction t_i (t_i is calculated as the ratio of the number of OS events observed at the time of the cut-off for the i^{th} interim analysis and the total number of OS events targeted for the final analysis.

For each comparison, the critical value $u(t_1)$ for the 1st interim analysis of OS is determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1).$$

Critical boundaries for the additional interim analyses and the final analysis of OS are calculated recursively as follows for each comparison

u(t2) is derived such that $\alpha(t_1) + P_0(Z_1 < u(t_1), Z_2 \geq u(t_2)) = \alpha(t_2)$,

u(t3) is derived such that $\alpha(t_2) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 \geq u(t_3)) = \alpha(t_3)$,

The boundary for the final efficacy analysis is derived such that

$$\alpha(t_3) + P_0(Z_1 < u(t_1), Z_2 < u(t_2), Z_3 < u(t_3), Z_F \geq u_F) = \alpha'$$

Where α' = 0.025 or 0.003 or 0.022 for the analysis of OS in patients unselected for DDR status as determined according to the testing strategy described in Section 5.1.1.

As described in Section 5.1.2, if the number of OS events in the final analysis deviates from the target number of OS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analyses and the actual correlation between the four test statistics Z_i (i=1, 2, 3) and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at α'.

8. REFERENCES

1. Amit O, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis and recommendation from a PhRMA working group. European Journal of Cancer 47:1772-1778, 2011.

2. Brookmeyer R, Crowley JJ. A confidence interval for the median survival time. Biometrics. 38: 29-41, 1982.

3. Burger RA, Brady MF, Bookman MA et al. 2011. Incorporation of bevacizumab in the primary treatment of ovarian cancer. N Engl J Med. 365(26):2473-83.

4. Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika; 26, 404-413.

5. EuroQoL Group. Prepared by vanReenen M and Janssen B. EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. Version 2.1. April 2015.

6. FACIT. NCCN/FACT Ovarian Symptom Index-18 (FOSI-18) Scoring Guidelines (Version 2) 2008.

7. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of Life Research. 20:1727-1736.

8. Jennison C, Turnbull BW. In: Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000.

9. Jensen SE, Rosenbloom SK, Beaumont JL, et al. A new index of priority symptoms in advanced ovarian cancer. Gynecol Oncol 2011;120(2):214-9.

10. Jensen SE, Kaiser K, Lacson L, et al. Content validity of the NCCN-FACT ovarian symptom index-18 (NFOSI-18). Gynecol Oncol 2015;136(2):317-22.

11. Kalbfleisch JD, Prentice, RL. Statistical Analysis of Failure Time Data, 2nd Edition. Hoboken, Wiley Interscience. 2011.

CC
I [redacted]

13. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 53: 457-81, 1958.

14. Rabin R, de Charro F. EQ-5D-5L: a measure of health status from the EuroQol Group. Ann Med. 2001 Jul; 33 (5): 337-43. Review.

15. Royston, P. and Parmar, M.K. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in medicine*, 2011; *30*(19), pp.2409-2421.

16. Rustin, G et.al. Definitions for Response and Progression in Ovarian cancer Clinical Trials Incorporating RECIST 1.1 and CA-125 Agreed by the Gynecological Cancer Intergroup (GCIG)., Int J Gynecol Cancer. 2011 Feb;21(2):419-23

17. Santner TJ, Snell MK. (1980). Small-sample confidence intervals for p1-p2 and p1/p2 in 2 x 2 contingency tables. *Journal of the American Statistical Association*, 75:386–394

18. Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 1980; 67:145-53.

19. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health 2000.

20. Uno H, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. J Clin Oncol 2014; 32.

21. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4): 183–9.

9. APPENDICES

Appendix 1. Immune-Related Adverse Events

The MedDRA PTs and clusters for irAEs are defined in the Safety Review Plan (SRP) for avelumab.

Immune-related AEs (irAEs) will be programmatically identified as outlined in Table 22. This case definition is hierarchical, ie, each step is only checked for patients and events that have already met the prior step.

Table 22. Case Definition for irAEs

Step	Selection Criteria	Additional Notes
1	Event selected based on a list of pre-specified MedDRA PTs within clusters. These are included in the SRP as Tier1 events (Immune-mediated xxxx). If AE matches the list then it is in for the next step	
2	AE onset during 1st study drug administration or anytime thereafter through 90 days after last dose of study treatment.	This is regardless of start of new anti-cancer drug therapy and regardless of TEAE classifications
3	Answer in the AE eCRF page to 'Was another treatment given because of the occurrence of the event' is 'YES'	
4	AE treated with corticosteroids or other immunosuppressant therapy. For endocrinopathies only: AE required hormone replacement	Look in the conmed pages for AE identifiers that match the AEs from Step 3. For each of such AEs if A) OR B) below are met then the AE is in for the next step A) conmed ATC code is in (H02A, H02B, D07, A01AC, S01BA, S01BB, L04AA, L04AB, L04AC, L04AD, L04AX, A07EA) and AE PT is in any of the irAE clusters. B) conmed ATC code is in (H03A, H03B) and AE PT is in one of the irAE clusters associated with "Immune-mediated endocrinopathies" C) conmed ATC code is A10A and AE PT is in the irAE cluster associated with "Immune-mediated endocrinopathies: Type I Diabetes Mellitus"

5	A) No clear etiology (other than immune mediated etiology)	A) From the AE eCRF page. Is the AE clearly related to an etiology other than immune-mediated etiology? Yes / No If answer is Yes, check all that apply: • Underlying malignancy / progressive disease. • Other medical conditions. • Prior or concomitant medications / procedures. • Other. Specify.
	B) Histopathology / biopsy consistent with immune-mediated event	B) From the AE eCRF page. B1) Was there a pathology /histology evaluation performed to investigate the AE? Y/N B2) If answer to the above is Yes, does the pathology/histology evaluation confirms an immune mediated mechanism for the AE? Y/N B3) If pathology / histology evaluation performed to investigate the AE, provide summary of relevant findings of the pathology /histology report. (Free Text)
	Event is in if [Answer to 5B1 and 5B2 is YES (regardless of answer to 5A)] OR [Answer to 5B1 is YES AND answer to 5B2 is NO AND answer to 5A is NO] OR [Answer to 5B1 is NO AND answer to 5A is NO]	

The data set associated with irAEs may be refined based on medical review. The final data set including any changes based on medical review (eg, addition of cases that are not selected by the programmatic algorithm) will be the basis of the irAE analyses.

Appendix 2. Infusion Related Reactions

For defining an AE as IRR the onset of the event in relation to the infusion of study drug and time to resolution of the event will be considered.

- All AEs identified by the MedDRA PT query describing signs and symptoms will be considered potential IRRs when onset is on the day of study drug infusion (during or after infusion) and the event resolved with end date within 2 days after onset.
- All AEs identified by the MedDRA PTs of Infusion related reaction, Drug hypersensitivity, Anaphylactic reaction, Hypersensitivity, Type 1 hypersensitivity, will be considered potential IRRs when onset is on the day of study drug infusion (during or after the infusion) or the day after the study drug infusion (irrespective of resolution date).

The list of MedDRA PTs for 'IRRs SIGNS and SYMPTOMS' and PTs 'IRRs CORE' are defined in the SRP for avelumab.

Infusion-related reactions (IRRs) will be programmatically identified as outlined in Table 23 or Table 24 and will be identified for IV drugs only.

Table 23. Case Definition for IRRs – IV Study Drugs Administered Alone Or In Combination With Non-IV Study Drugs

Condition	Selection criterion
If AE meets [1 AND 2] OR [3 AND (4A OR 4B)] then AE is classified as an IRR	
1	PT is included in the 'IRRs SIGNS and SYMPTOMS' list
2	• AE onset date = date of infusion of study drug AND • AE timing related to study drug ('DURING', 'AFTER') AND • AE outcome in ('RECOVERED/RESOLVED', 'RECOVERED/RESOLVED WITH SEQUELAE', 'RECOVERING/RESOLVING') AND • AE end date – AE onset date <=2
3	PT is included in the 'IRRs CORE' list
4A	• AE onset date = date of infusion of study drug AND • AE timing related to study drug in ('DURING', 'AFTER')
4B	AE onset on the day after infusion

Table 24. Case Definition for IRRs – IV Study Drugs Administered in Combination (eg, Doublets or Triplets)

Condition	Selection criterion
IRR can be associated with the first IV drug and/or subsequent IV drugs that are administered in combination. Without loss of generality assume triplet IV with D_1 administered first then D_2 then D_3. The IV study drug or drugs associated with the IRR need to be identified in the analysis data set to enable subsequent analysis. The following are not sequential and an AE can be classified as an IRR associated with multiple D_J from one or more of I, II, III, IV, V below: I - If the AE meets [1 AND 2A1] for a D_J then the AE is classified as an IRR associated with the D_J that meets the 2A1 criterion II - If the AE meets [1 AND 2A2] for a D_J then the AE is classified as an IRR associated with the D_J and associated with D_{J+1} that meets the 2A2 criterion III - If the AE meets [3 AND 4B] for any D_J then the AE is classified as an IRR associated with all D_J that meet the 4B criterion. IV- If the AE meets [3 AND 4A1] for a D_J then the AE is classified as an IRR associated with the D_J that meets the 4A1 criterion V- If the AE meets [3 AND 4A2] for a D_J then the AE is classified as an IRR associated with the D_J and associated with D_{J+1} that meets the 4A2 criterion	
1	PT is included in the 'IRRs SIGNS and SYMPTOMS' list
2A1	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • [AE timing related to study drug D_{J+1} ('BEFORE') OR AE onset date < date of infusion of study drug D_{J+1}]AND • AE outcome in ('RECOVERED/RESOLVED', 'RECOVERED/RESOLVED WITH SEQUELAE', 'RECOVERING/RESOLVING') AND • AE end date – AE onset date <=2
2A2	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • AE timing related to study drug D_{J+1} ('DURING', 'AFTER') AND • AE outcome in ('RECOVERED/RESOLVED', 'RECOVERED/RESOLVED WITH SEQUELAE', 'RECOVERING/RESOLVING') AND • AE end date – AE onset date <=2
3	PT is included in the 'IRRs CORE' list
4A1	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • [AE timing related to study drug D_{J+1} ('BEFORE') OR AE onset date < date of infusion of study drug D_{J+1}]

4A2	• AE onset date = date of infusion of study drug D_J AND • AE timing related to study drug D_J ('DURING', 'AFTER') AND • AE timing related to study drug D_{J+1} ('DURING', 'AFTER')
4B	AE onset on the day after infusion of study drug D_J