Official Title: A Phase III, Multicenter, Randomized, Study of Atezolizumab Versus Placebo Administered in Combination With Paclitaxel, Carboplatin, and Bevacizumab to Patients With Newly-Diagnosed Stage III or Stage IV Ovarian, Fallopian Tube, or Primary Peritoneal Cancer

NCT Number: NCT03038100

Document Date: SAP Version 3: 06-Dec-2021

STATISTICAL ANALYSIS PLAN

TITLE: **A PHASE III, MULTICENTER, RANDOMIZED, STUDY OF ATEZOLIZUMAB VERSUS PLACEBO ADMINISTERED IN COMBINATION WITH PACLITAXEL, CARBOPLATIN, AND BEVACIZUMAB TO PATIENTS WITH NEWLY-DIAGNOSED STAGE III OR STAGE IV OVARIAN, FALLOPIAN TUBE, OR PRIMARY PERITONEAL CANCER**

PROTOCOL NUMBER: YO39523 (GOG-3015, ENGOT–ov39)

STUDY DRUG: Atezolizumab (RO5541267)

VERSION NUMBER: 3

IND NUMBER: 130637

EUDRACT NUMBER: 2016-003472-52

SPONSOR: F. Hoffmann-La Roche Ltd

PLAN PREPARED BY: [redacted], M.S.

DATE FINAL: Version 1: 16 September 2019

DATE AMENDED Version 2: 14 November 2019
Version 3: See electronic date stamp below

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Date and Time(UTC)	Reason for Signing	Name
06-Dec-2021 16:03:42	Company Signatory	[redacted]

CONFIDENTIAL

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

This Statistical Analysis Plan (SAP) Version 3 for Study YO39523 has been amended to incorporate a following change:

- In Section 2.5, additional description of End of Study has been included

Additional minor changes have been made to improve clarity and consistency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. BACKGROUND

This Statistical Analysis Plan (SAP) describes the analyses that are planned to be performed for the Clinical Study Report (CSR) of Study YO39523 (IMagyn050).

It is anticipated that positive results from Study YO39523 will support the submission of filing applications globally for the use of atezolizumab in combination with paclitaxel, carboplatin, and bevacizumab in patients with newly diagnosed Stage III or Stage IV ovarian, fallopian tube, or primary peritoneal cancer. For purposes of registration, the analyses outlined in this SAP will supersede those specified in the Protocol.

Please refer to Section 1 of the Protocol for more details about atezolizumab and the study rationale of Study YO39523.

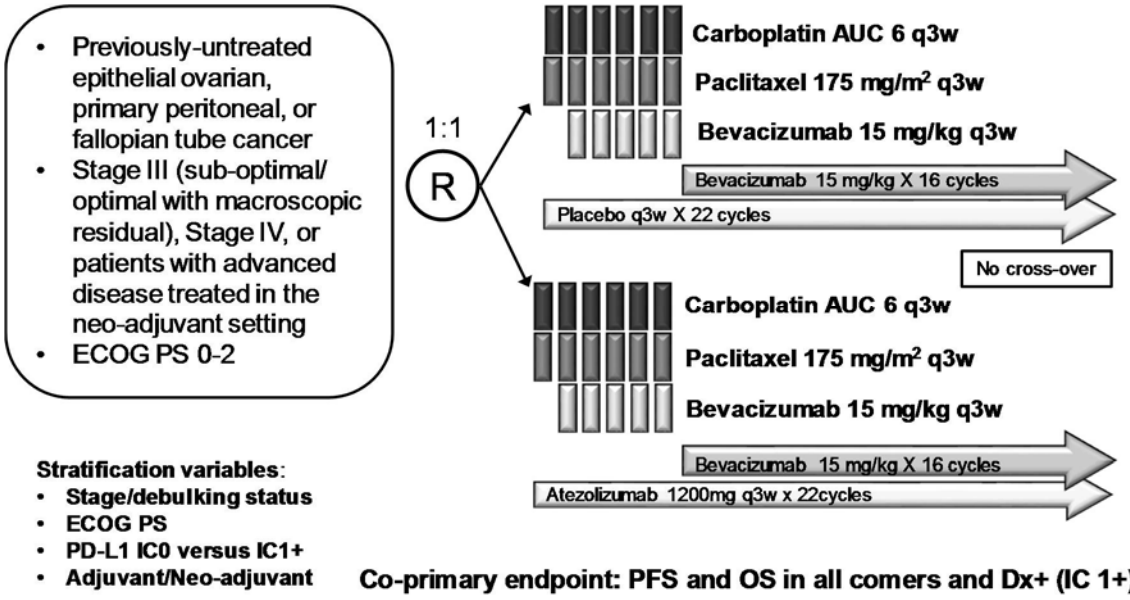
2. STUDY DESIGN

This is a Phase III double-blind, two-arm, randomized study designed to evaluate the efficacy and safety of atezolizumab (1200 mg) administered with paclitaxel (175 mg/m2)+carboplatin (area under the concentration–time curve at 6 [AUC 6])+bevacizumab (15 mg/kg) compared with placebo+paclitaxel+carboplatin+bevacizumab in patients with Stage III or Stage IV ovarian, fallopian tube, or primary peritoneal cancer with macroscopic residual disease postoperatively (i.e., after primary tumor reductive surgery) or who will undergo neoadjuvant therapy followed by interval surgery. Approximately 1300 patients will be randomized, in a 1:1 ratio, to one of the two treatment arms. After the enrollment of global cohort finishes, additional Chinese patients from mainland China might continue to enroll into the China extension cohort. Chinese patients enrolled in global cohort combined with patients in China extension cohort will consist of China subgroup.

More details about the study design could be found in Appendix 1, Protocol Synopsis.

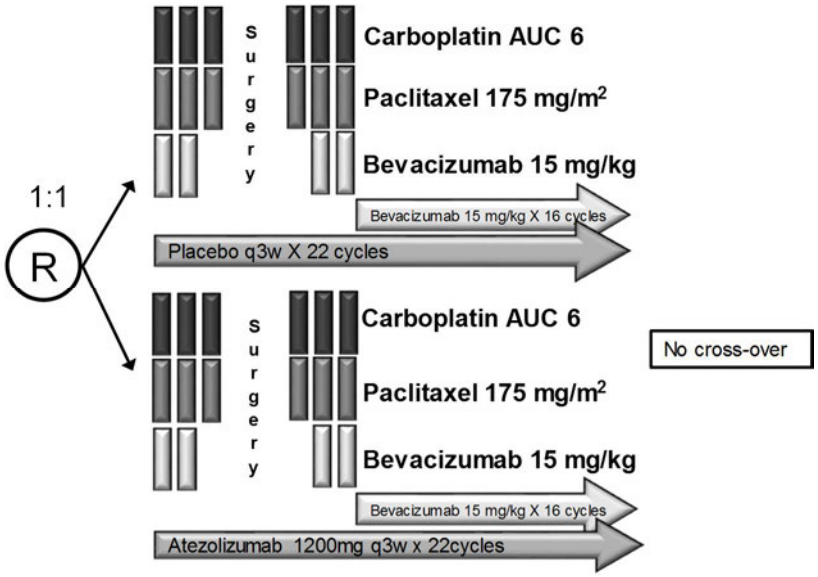
Figure 1 and Figure 2 illustrate the study design.

Figure 1 Study Schema for Patients in the Primary Tumor-Reductive Surgery Group



AUC = area under the concentration–time curve; Dx = diagnosis; ECOG PS = Eastern Cooperative Oncology Group performance status; OS = overall survival; PFS = progression-free survival; PD-L1 = programmed death–ligand 1; q3w = every 3 weeks.

Figure 2 Study Schema for Patients in the Neoadjuvant Chemotherapy Group



AUC = area under the concentration–time curve; q3w = every 3 weeks.

Patients will undergo a mandatory tumor biopsy sample collection, if clinically feasible, at the time of first evidence of radiographic disease progression according to Response Evaluation Criteria in Solid Tumors, Version 1.1 (RECIST v1.1). Cytology from ascites,

pleural effusion, and fine needle aspiration (FNA) is not adequate. These samples will be retrospectively analyzed to evaluate and/or characterize pseudoprogression caused by immune cells (ICs) from true progression. In addition, tumor tissue biomarkers related to resistance, disease progression, and clinical benefit of atezolizumab will be analyzed.

2.1 STUDY OBJECTIVES

Study objectives and corresponding endpoints for the study are outlined below (Table 1).

Table 1 Objectives and Corresponding Endpoints

Primary Efficacy Objective	Corresponding Endpoints
• To evaluate the efficacy of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab among all patients and in those patients with PD-L1–positive tumors	• Investigator-assessed PFS, defined as the time from randomization to the occurrence of disease progression, as determined by the investigator from tumor assessments per RECIST v1.1, or death from any cause during the study, whichever occurs first • OS, defined as the time from randomization to death from any cause
Secondary Efficacy Objectives	**Corresponding Endpoints**
Among patients with measurable residual disease in the primary surgery group*:	
• To evaluate the efficacy of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab among all patients and in those patients with PD-L1–positive tumors	• OR, defined for primary surgery patients with measurable disease at baseline who achieved a documented unconfirmed response [i.e., either a partial response (PR) or a complete response (CR)] on the basis of investigator assessment using RECIST v1.1.
• To evaluate the duration of efficacy observed with atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab among all patients and in those patients with PD-L1–positive tumors	• DOR, is defined for primary surgery patients who had an objective response as the time from the first occurrence of a documented unconfirmed response (CR or PR) to the date of disease progression on the basis of investigator assessment using RECIST v1.1 or death from any cause, whichever occurs first.

Table 1 Objectives and Corresponding Endpoints (cont.)

Secondary Efficacy Objectives (cont.)	Corresponding Endpoints (cont.)
Among the patients in the neoadjuvant group:	
• To determine the impact of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab on patient-reported abdominal symptoms of OC, as measured by two items from the abdominal/GI symptom scale of the EORTC QLQ-OV28	• Proportion of patients with clinically-meaningful improvement in patient-reported abdominal pain or bloating, defined as a ≥ 10-point decrease from the baseline score on either of the two items on the EORTC QLQ-OV28 abdominal/GI symptom scale (items 31 and 32)
Among the patients in the primary surgery and neoadjuvant group:	
• To evaluate PROs of HRQoL associated with atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab, as measured by the functional and global health status (GHS)/QoL scales of the EORTC QLQ-C30	• In the neoadjuvant group: Proportion of patients with clinically-meaningful improvement in patient-reported function and HRQoL, defined as a ≥ 10-point increase from the baseline score on each of the functional (physical, role, emotional, and social) and GHS/QoL scales of the EORTC QLQ-C30 • In the primary surgery group: Proportion of patients with clinical improvement, remaining stable, or deterioration in patient-reported function and HRQoL, defined as a ≥ 10-point increase, changes within 10 points, and a ≥ 10-point decrease, respectively, from the baseline score on each of the functional (physical, role, emotional, and social) and GHS/QoL scales of the EORTC QLQ-C30
Safety Objective	**Corresponding Endpoints**
• To evaluate the safety and tolerability of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab	• Occurrence and severity of adverse events, with severity determined in accordance to NCI CTCAE v4.0 • Change from baseline in targeted vital signs • Change from baseline in targeted clinical laboratory test results
Pharmacokinetic Objective	**Corresponding Endpoint**
• To characterize the pharmacokinetics of atezolizumab when administered in combination with paclitaxel + carboplatin + bevacizumab	• Minimum and maximum serum concentration of atezolizumab

Table 1 Objectives and Corresponding Endpoints (cont.)

Exploratory Objectives	Corresponding Endpoints
• Among neoadjuvant patients only: To evaluate PCR status and its association with clinical outcomes after administration of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab	• Among patients who undergo neoadjuvant therapy prior to interval surgery, PCR status is defined as the clinical amount and histologic characteristics of residual disease assessed at the time of interval cytoreductive surgery
• To evaluate the efficacy of atezolizumab versus placebo administered in combination with paclitaxel + carboplatin + bevacizumab	• The OS rate at 3 years after randomization
• To evaluate PROs of disease and/or treatment-related symptoms associated with atezolizumab versus placebo administered in combination with paclitaxel + carboplatin + bevacizumab, as measured by the EORTC QLQ-C30 and QLQ-OV28	• Mean and mean changes from the baseline score in disease and/or treatment-related symptoms by cycle and between treatment arms as assessed by all symptom items and/or scales of the EORTC QLQ-C30 and QLQ-OV28
• To evaluate any treatment burden patients may experience in association with the addition of atezolizumab to paclitaxel + carboplatin + bevacizumab compared with placebo + paclitaxel + carboplatin + bevacizumab, as measured by a single item (from GP5: "I am bothered by side effects of treatment") from the physical wellbeing subscale of the FACT-G Quality of Life instrument	• Proportion of patients reporting each response option at each assessment timepoint by treatment arm for item GP5 from the FACT-G

Table 1 Objectives and Corresponding Endpoints (cont.)

Immunogenicity Objective	Corresponding Endpoint
• To evaluate the immune response from patients who were administered atezolizumab	• The incidence of ADAs against atezolizumab during treatment with atezolizumab administered in combination with paclitaxel + carboplatin + bevacizumab relative to the incidence of ADAs at the baseline
Exploratory Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate the potential effects of ADAs	• The relationship between ADA status and pharmacokinetics, safety, and efficacy

ADA = anti-drug antibody; CR = complete response; DOR = duration of response; EORTC = European Organisation for Research and Treatment of Cancer; EQ-5D-5L = EuroQoL 5 Dimension, 5 Level Questionnaire; FACT-G = Functional Assessment of Cancer Therapy-General; GI = gastrointestinal; HRQoL = health-related quality of life; IHC = immunohistochemistry; PCR = pathologic and clinical response; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events; NGS = next-generation sequencing; OC = ovarian cancer; OR = objective response; OS = overall survival; PFS = progression-free survival; PR = partial response; PRO =patient-report outcome; QLQ-C30 = Quality of Life Questionnaire Core 30; QLQ-OV28 = Quality of Life Questionnaire Ovarian Cancer Module 28; RECIST v1.1 = Response Evaluation Criteria in Solid Tumors, Version 1.1.

2.2 PROTOCOL SYNOPSIS

The Protocol Synopsis is in Appendix 1. For additional details, see the Schedules of Assessments in Appendix 2, Appendix 3, and Appendix 4.

2.2.1 Endpoints

See the Protocol Synopsis in Appendix 1 for a description of the primary outcome measures.

2.3 TYPE I ERROR CONTROL

The type I error (alpha [α]) for this study is 0.05 (two-sided). Type I error will be controlled for the following efficacy endpoints in intent-to-treat (ITT) and programmed death–ligand 1 (PD–L1)–positive populations:

- Co-primary efficacy endpoint of investigator-assessed progression-free survival (PFS) by RECIST v1.1 (ITT and PD-L1-positive populations).
- Co-primary efficacy endpoint of overall survival (OS) (ITT and PD-L1-positive populations).

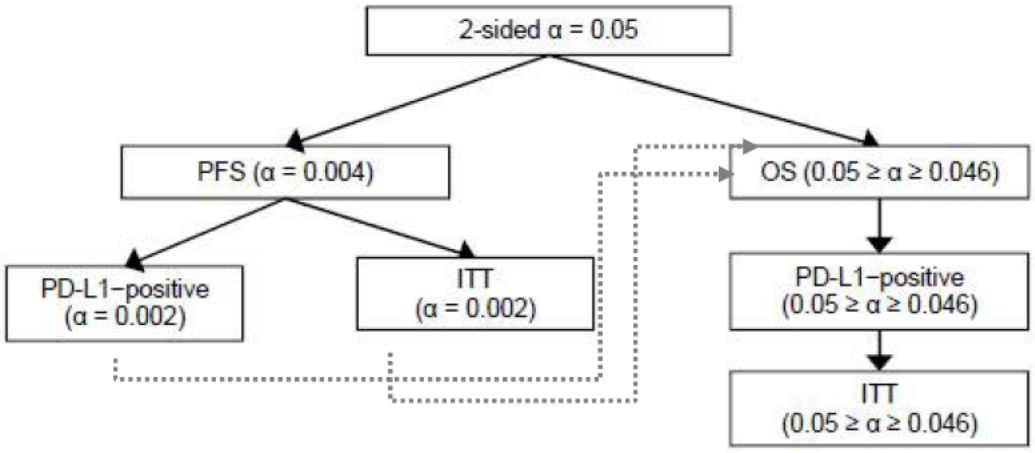
In IMagyn050, the control of type I error rate at 5% is performed using a combination of alpha split and hierarchical testing procedures. The alpha of this study is fully controlled by the closed test procedure (Dmitrienko et al. 2013).

Based on the potential scenarios regarding the alpha level for the final analysis of OS, the boundaries for statistical significance and the expected timing of each interim analyses

and the final analysis were determined based on the Lan-DeMets implementation of the O'Brien-Fleming use function (DeMets et al. 1994). Interim analyses are described in Section 4.9.1.

The alpha level for OS was determined at the time of the primary PFS analysis. The total alpha ($\alpha = 0.05$) was divided among the two co-primary endpoints composed of PFS tested in the intent-to-treat (ITT) and PD-L1-positive population ($\alpha_{PFS} = 0.004$) and in the ITT (α_{OS}=0.046), and maintained a fixed sequence for the testing (i.e., PFS in ITT and PD-L1-positive tested first) as per Figure 3.

Figure 3 Overall of Type I Error Control



ITT = intent-to-treat; PD-L1 = programmed death–ligand 1; PFS = progression-free survival; OS = overall survival.

At the primary PFS analysis, as the testing sequence progressed, a successful test for PFS in ITT and PD-L1-positive populations preserves its assigned alpha ($\alpha_{PFS}=0.002$ in each population) as "saved" (unused) alpha that is passed along to the next tests in the sequence, as is the case for the sequential method:

- As the PFS in the ITT population is significant at level $\alpha_{PFS}=0.002$, this alpha is unused and is passed to the second endpoint test as an additional alpha of 0.002 for OS in PD–L1 positive.
- As the PFS in the PD-L1-positive population is significant at level $\alpha_{PFS}=0.002$, this alpha is unused and is passed to the second endpoint test as an additional alpha of 0.002 for OS in PD–L1 positive.

The passed-along alpha related to PFS from the ITT and PD-L1-positive populations (i.e., 0.002 for each population) is added to the prospectively assigned alpha of OS endpoint ($\alpha_{OS}=0.046$) and the summed alpha ($\alpha_{OS} \geq 0.046$) is used for testing that endpoint.

2.4 DETERMINATION OF SAMPLE SIZE

Approximately 1300 patients will be randomized 1:1 to the two treatment arms.

The sample size of the study is determined by the number of patient deaths required to demonstrate efficacy in terms of OS in the PD-L1–positive subgroup and the ITT population. To detect an improvement in OS with the use of a log-rank test at a two-sided significance level of 0.046, approximately 311 deaths in the PD-L1–positive subgroup will be required to achieve 81% power with a target HR of 0.72, and approximately 534 deaths in the ITT population to achieve 80% power with a target HR of 0.78. The numbers of patient deaths needed for the OS analyses in the PD-L1–positive

subgroup and the ITT population are listed in Appendix 4 for other significance levels that could be assigned to the OS analyses.

2.4.1 Assumptions

The calculation of the sample size and estimates of the analysis timelines are based on the following assumptions:

- PFS and OS are exponentially distributed.
- The median duration of PFS in the control arm is 18 months.
- The median duration of OS in the control arm is 43 months.
- The prevalence of PD–L1–positive (IC1/2/3) patients is 60%.
- The two interim and final analyses of OS use the Lan-DeMets alpha spending function to approximate the O'Brien-Fleming boundary.
- The dropout rate is 5% over 12 months for PFS and OS.
- The recruitment of 1300 patients will take place over 25 months.

2.4.2 Co-Primary Endpoint: Progression-Free Survival

The primary analysis of PFS will take place when approximately 601 PFS events in the ITT and 347 PFS events in the PD–L1–positive subgroup have occurred, which is expected at approximately 36 months after the first patient is enrolled in the study. This provides 90% power to detect a PFS improvement of HR = 0.7 in the ITT population, and 91% power in the PD–L1–positive subgroup with HR = 0.62, at a two-sided significance level of 0.002.

2.4.3 Co-Primary Endpoint: Overall Survival

The numbers of patient deaths needed for the OS analyses in the PD–L1–positive subgroup and the ITT population are listed in Section 4.9.1 for possible significance levels that could be assigned to the OS analyses. For example, to detect an improvement in OS with the use of a log-rank test at a two-sided significance level of 0.046, approximately 311 deaths in the PD-L1–positive subgroup will be required to achieve 81% power with a target HR of 0.72, and approximately 534 deaths in the ITT population to achieve 80% power with a target HR of 0.78.

Two interim analyses of OS will be performed on patients who are in the ITT and PD-L1–positive populations. The timing of the two interim analyses and the final analysis for OS depends on the results from the primary analysis of the co-primary endpoint of PFS, which is listed in Section 4.9.1 and includes the pre-specified boundaries for the different scenarios.

2.4.4 Sample Size for the China Extension Period

Refer to Section 6.9 of the Protocol for more details.

2.5 END OF STUDY

The study will continue until the approximate pre-planned numbers of deaths among the PD-L1-positive patients and the ITT population have been observed (see Appendix 4 of the study protocol). In addition, the Sponsor has the right to close the study at any time if futility is observed (e.g., based on the predicted probability of success at the subsequent analysis).

3. STUDY CONDUCT

3.1 RANDOMIZATION

Randomization occurs in a 1:1 ratio using a permuted-block randomization method into one of two treatment arms: atezolizumab + paclitaxel + carboplatin + bevacizumab or placebo + paclitaxel + carboplatin + bevacizumab. The randomization scheme is designed to ensure that an approximately equal number of patients will be enrolled in each treatment arm within the baseline characteristics of the following stratification factors:

- Stage and/or residual disease status (Stage III vs. Stage IV)
- Eastern Cooperative Oncology Group (ECOG) performance status (0 vs. 1 or 2)
- Tumor PD-L1 status (IC0 vs. IC1/2/3)
- Treatment strategy (adjuvant vs. neoadjuvant)

3.2 DATA MONITORING

An independent Data Monitoring Committee (iDMC) will monitor safety and study conduct on a periodic basis. Members of the iDMC will be external to the Sponsor and will follow a charter that outlines their roles and responsibilities. The iDMC will meet approximately every 6 months from the point of FPI to review unblinded safety and study conduct data prepared by an independent Data Coordinating Center (iDCC). The safety data will include demographic data, adverse events (AEs), serious adverse events (SAEs), and relevant laboratory data.

Following each data review, the iDMC will provide recommendations to the Sponsor as to whether the study should continue or be amended, or whether the study should be stopped on the basis of safety (i.e., evidence of harm). The Sponsor's Data Review Board (DRB; a group consisting of employees of the Sponsor empowered to make critical decisions) will make a decision on the basis of the iDMC's recommendations. The DRB of the Sponsor will either accept or reject this recommendation. The final decision will rest with the Sponsor.

Any outcomes of these safety reviews that affect study conduct will be communicated in a timely manner to the investigators for notification of the Institutional Review Boards/Ethics Committees (IRBs/ECs).

Members of the iDMC will be external to the Sponsor and will follow a separate iDMC Charter that outlines their roles and responsibilities, as well as a detailed monitoring plan.

In addition, iDMC will make assessment respect to the PFS audit strategy, and send their recommendation to the Sponsor without revealing any numeric value (see Appendix 5).

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

The analysis populations are defined as follows:

- The ITT population is defined as all randomized patients, whether or not the assigned study treatment was received.
- The PD-L1-positive subpopulation is defined as patients in the ITT population whose PD-L1 status is IC1/2/3 at the time of randomization.
- The ORR-evaluable population is defined as primary surgery patients in the ITT population with measurable disease at baseline.
- The PD-L1−ORR-evaluable population is defined as primary surgery patients in the PD-L1-positive subpopulation with measurable disease at baseline.
- The duration of response (DOR)-evaluable population is defined as patients with an objective response.
- The patient-reported outcome (PRO)-evaluable population is defined as patients in the ITT population with a baseline and ≥1 post-baseline PRO assessment.
- The safety-evaluable population is defined as patients who received any amount of any study drug.
- The pharmacokinetic (PK)-evaluable population is defined as all patients who received any dose of study medication and who have at least one post-baseline PK sample available.

For all efficacy analyses, patients will be grouped according to the treatment assigned at randomization.

For safety analyses, patients will be grouped according to whether any amount of atezolizumab was received, including cases in which atezolizumab was received in error.

4.2 ANALYSIS OF STUDY CONDUCT

Study enrollment, patient disposition, reasons for discontinuation from the study treatment and reason for study termination will be summarized for all patients in the ITT population.

Protocol deviations, including violations of inclusion/exclusion criteria and deviations during study conduct will be reported and summarized.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographics, baseline disease characteristics, and ovarian cancer history will be compared between both treatment arms. Descriptive baseline summaries of continuous data will present the group mean, standard deviation, median, ranges and inter-quartile

ranges, as appropriate. Descriptive baseline summaries of discrete data will present the category counts as frequencies and percentages.

The baseline value of any efficacy variable will be defined as the last available value recorded prior to randomization.

The baseline value of any non-efficacy variable will be defined as the last available value recorded prior to the first administration of study medication.

Previous and concomitant cancer therapy will also be summarized, including radiotherapy and surgery, as well as subsequent anti-cancer therapy. Previous and concurrent diseases and medications will also be summarized.

4.4 EFFICACY ANALYSIS

All efficacy analyses will be performed on both the ITT population and PD-L1-positive subpopulation.

4.4.1 Primary Efficacy Endpoint

4.4.1.1 Progression-Free Survival

Progression-free survival (PFS) is defined as the time from randomization to the occurrence of disease progression, as determined by investigators from tumor assessments, per RECIST v1.1, or death from any cause, whichever occurs first. Patients who have not experienced disease progression or death at the time of analysis will be censored at the time of the last tumor assessment. Patients with no post-baseline tumor assessment or death will be censored on the date of randomization+1 day.

Progression-free survival (PFS) will be analyzed simultaneously in the ITT and PD-L1-positive subgroup (see Section 2.2.1). The following analyses will be performed for both PFS endpoints described above:

- Treatment comparisons will be based on the stratified log-rank test. The stratification factors will be those used for randomization (see Section 3.1) and will be obtained from the IxRS. Results from an un-stratified analysis and a stratified analysis using eCRF-collected stratification factors will also be provided.
- The HR will be estimated using a stratified Cox regression model with the same stratification variables used for the stratified log-rank test, and the 95% CI for the HR will be provided. Results from an un-stratified analysis will also be provided.
- Kaplan-Meier methodology will be used to estimate median PFS for each treatment arm and to construct survival curves for each treatment arm. The Brookmeyer-Crowley methodology will be used to construct the 95% CI for the median PFS for each treatment arm (Brookmeyer and Crowley 1982).

4.4.1.2 Overall Survival

Overall survival is defined as the time from the date of randomization to the date of death due to any cause. Patients who are not reported as having died at the time of analysis

will be censored at the date when they were last known to be alive. Patients who do not have post-baseline information will be censored at the date of randomization + 1 day. Testing of OS is outlined in Section 2.3 and analysis of OS is performed analogously to PFS in the section above. OS is hierarchically tested in the PD-L1-positive and ITT population (see Section 2.4).

4.5 INDEPENDENT REVIEW FACILITY

The imaging data used for tumor assessment will be collected by the Sponsor to enable centralized, independent review of response endpoints by an Independent Review Committee (IRC). An independent imaging group will be used to evaluate tumor assessments for determination of progression and response rate according to RECIST v1.1 for the primary analysis. Imaging studies (computed tomography [CT]/magnetic resonance imaging [MRI]/bone scans) will be acquired according to a standard protocol and will be forwarded to the independent reviewers. Investigator tumor assessments will not be reconciled with the IRC tumor assessments. Further details will be included in the IRC Charter. Details of imaging handling procedures are also described in a separate laboratory manual. An independent Data Monitoring Committee (iDMC) will be in charge to evaluate and assess if the level of bias in local evaluations is unacceptable OR if a full independent review is unnecessary (see Section 3.2).

4.5.1 Secondary Efficacy Endpoints

4.5.1.1 Objective Response Rate (ORR)

An objective response is defined for primary surgery patients with measurable disease at baseline who achieved a documented unconfirmed response [i.e., either a partial response (PR) or a complete response (CR)] on the basis of investigator assessment using RECIST v1.1. Patients not meeting this criterion, including patients without any post-baseline tumor assessment, will be considered as non-responders. Objective response rate is defined as the proportion of primary surgery patients with measurable disease at baseline who have an objective response.

Objective response rate will be compared between treatment arms using the stratified Cochran-Mantel-Haenszel test. Primary surgery and neoadjuvant patients (see Section 4.5.4.1) will be evaluated separately. The stratification factors will be the same as those described for the analysis of the primary endpoint of PFS. The difference in ORR between treatment arms will be calculated, and its 95% CI will be calculated using the normal approximation to the binomial distribution. An estimate of ORR will be calculated for each treatment arm, and its 95% CI will be calculated using the Clopper-Pearson method. ORR is simultaneously assessed in the ITT and PD-L1 populations with measurable disease at baseline.

4.5.1.2 Duration of Objective Response

Duration of response is defined for primary surgery patients who had an objective response as the time from the first occurrence of a documented unconfirmed response (CR or PR) to the date of disease progression on the basis of investigator assessment

using RECIST v1.1 or death from any cause, whichever occurs first. Data for patients who have not experienced disease progression or death will be censored at the last tumor assessment date. If no tumor assessments were performed after the date of the first occurrence of the objective response (CR or PR), patient will be censored at the date of the first occurrence of the objective response + 1 day.

The analysis of DOR is based on a non-randomized subset of patients (those who achieved an unconfirmed response); therefore, no formal hypothesis testing will be performed. Comparisons between treatment arms will be made for descriptive purposes only. The methodologies described for the analysis of PFS will be used for the analysis of DOR except that the analysis will not be stratified.

4.5.1.3 Patient-Reported Outcomes Disease Symptoms, Function and Health-Related Quality of Life EORTC Data

Patient-Reported Abdominal Pain or Bloating and EORTC QLQ-OV28: For patients in the neoadjuvant therapy group, the proportion of patients in each arm who report a clinically-meaningful improvement in patient-reported abdominal pain or bloating, defined as a ≥ 10-point decrease from the baseline score on each of two items from the European Organisation for Research and Treatment of Cancer Quality of Life Questionn aire Ovarian Cancer Module 28 (EORTC QLQ-OV28) abdominal/gastrointestinal (GI) symptom scale (items 31 and 32), will be summarized at each post-baseline timepoint by treatment arm.

The Cochran-Mantel-Haenszel test, stratified by stage and/or residual disease status (Stage III vs. Stage IV), ECOG performance status (0 vs. 1 or 2), and tumor PD-L1 status (IC0 vs. IC1/2/3) will be used to compare the proportion of patients who report a clinically-meaningful improvement in patient-reported abdominal pain or bloating at 9 weeks after neoadjuvant therapy between the two treatment arms. The difference in proportions will be provided, with its 95% CI, with the use of the Hauck-Anderson method.

Pre-specified subgroup analysis will also be performed in patients with ascites at baseline (who typically have significantly impaired HRQoL) and in patients with sufficient symptoms at baseline to allow detection of a 10-point improvement in a given symptom score.

The definition of improvement in patient-reported abdominal pain or bloating (i.e., a ≥ 10-point decrease from the baseline score in QLQ-OV28 abdominal symptom items) is based on the standard analysis method for the EORTC QLQ-C30 that deems a score change of 10 points on any item or scale to be clinically meaningful (Osoba et al. 1998; Fayers 2001a; Osoba 2002; Osoba et al. 2005; Brundage et al. 2007; Luckett et al. 2010; Cocks et al. 2011). Although the clinical meaningfulness of a 10-point change was established based on the EORTC QLQ-C30, the disease-specific modules, including the QLQ-OV28, were designed on the same structure using the same rating scale and are, therefore, applicable in this context. Additionally, other OC studies have used the 10-point minimally important difference (MID) threshold for the QLQ-OV28,

demonstrating that a change of this magnitude is significant to patients with OC while setting a precedent for its use and supporting its utility in this context (Richter et al. 2012; Brotto et al. 2016; Fagotti et al. 2016).

A sensitivity analysis will be performed to evaluate the robustness of the published standard threshold for meaningful change of 10-points with the use of the raw data for the abdominal pain and bloating items of the EORTC QLQ-OV28. The proportion of patients in each arm reporting a 1-category decrease on each of the 4-point symptom scales of the EORTC QLQ-OV28 abdominal symptom items (items 31 and 32), will be summarized at each post-baseline timepoint by treatment arm. All analyses of the abdominal symptoms single item data that involve the 10-point MID will be replicated with this alternate MID threshold.

All EORTC QLQ-OV28 data will be scored according to the EORTC scoring manual (Fayers et al. 2001b). PRO completion, compliance rates, and reasons for missing data will be summarized at each timepoint by treatment arm.

Patient-Reported Function and HRQoL EORTC QLQ-C30: For patients in the neoadjuvant therapy subgroup, the proportion of patients in each arm who report a clinically-meaningful improvement in patient-reported function and HRQoL, defined as a ≥10-point increase from the baseline score on each of the functional (physical, role, emotional, and social) and GHS/QoL scales of the EORTC QLQ-C30, will be summarized at each post-baseline timepoint by treatment arm, and compared between treatment arms with the use of the stratified Cochran-Mantel-Haenszel test specified above.

For patients in the primary surgery group, the proportion of patients in each arm who improve, remain stable or deteriorate in patient-reported functions and GHS/QoL, defined as a $\geq$10-point increase, changes within 10 points, and a $\geq$10-point decrease, respectively, from the baseline score on each of the functional (physical, role, emotional, and social) and GHS/QoL scales of the EORTC QLQ-C30 will be summarized at each post-baseline timepoint by treatment arm, and compared between treatment arms with the use of the stratified Cochran-Mantel-Haenszel test specified above.

The EORTC QLQ-C30 and QLQ-OV28 will be scored according to the EORTC scoring manual (Fayers et al. 2001b). Per Fayers et al. (2001b), in the event that incomplete data exists, if the scale has more than 50% of the constituent items completed, a pro-rated score will be computed. For subscales with less than 50% of the items completed, the subscale will be considered missing. PRO completion, compliance rates, and reasons for missing data will be summarized at each timepoint by treatment arm and surgery status.

4.5.2 Sensitivity Analyses

Statistical methodologies that are analogous to those methodologies used in the primary analysis of PFS and OS as specified in Section 4.4.1.1 will be applied for sensitivity analyses.

4.5.2.1 Sensitivity Analyses of Progression-Free Survival

Censoring for Non-Protocol Therapy: Non-protocol therapy is defined as any anti-cancer therapy other than study treatment that typically is the subsequent line of therapy. The impact of NPT on the co-primary endpoint of investigator-assessed PFS will be evaluated. A sensitivity analysis will be performed in which data for patients who received NPT will be censored at the last tumor assessment date before the patient received NPT.

PFS by IRC: An analysis of PFS on the basis of the IRC assessments will be performed using the same methodology as specified for PFS on the basis of investigator assessment.

Missing Assessments: A sensitivity analysis will be performed when disease progression occurs after two or more consecutive missed tumor assessments, these events will not be counted; rather, the patient will be censored at the patient's last tumor assessment prior to the missing assessments. If disease progression occurs after one missed tumor assessments, the event will be counted at the respective event date.

Additional sensitivity analyses may be considered if appropriate.

4.5.2.2 Sensitivity Analyses of Overall Survival

Accounting for Second-Line Immunotherapy Use: Quickly evolving development of checkpoint inhibitors may lead to increased PD-1/PD-L1 treatment options for patients in the second-line triple-negative breast cancer, either via trial participation or newly approved medicines in this class. Second-line usage of such inhibitors by patients progressing on this first-line trial could result in biased estimate of the treatment effect on OS. To account for this possibility of bias the following sensitivity analyses will be conducted.

Censoring for Treatment Switching: Treatment switching is defined as any checkpoint inhibitor therapy other than study treatment as subsequent line of therapy. Censoring for treatment switching will be applied to OS, analogue to censoring for NPT for PFS, see above.

Additional sensitivity analyses may be considered if appropriate.

4.5.3 Subgroup Analyses

To assess the consistency of study results in the subgroups defined by demographic baseline characteristics, biomarker status, PFS and OS in these subgroups will be examined. Summaries of PFS and OS, including unstratified HRs estimated with the use of Cox proportional hazards models and Kaplan-Meier estimates of the median, will be produced separately for each level of the categorical variables.

4.5.4 Exploratory Analyses

Analysis of ORR in Patients of Neoadjuvant Group: An objective response among neoadjuvant patients is defined for those patients with measurable disease at baseline who achieved a documented unconfirmed response (i.e., either a partial response [PR] or a complete response [CR]) on the basis of investigator assessment using RECIST v1.1 prior to interval cytoreductive surgery. Patients not meeting this criterion, including patients without any post-baseline tumor assessment, will be considered as non-responders. Objective response rate (ORR) among neoadjuvant patients is defined as the proportion of neoadjuvant patients with measurable disease at baseline who have an objective response, see the definition of Section 4.4.2.1, before interval surgery.

Pathologic and Clinical Response: The pathologic and clinical response (PCR) status is defined as the extent of residual disease assessed at the time of interval cytoreductive surgery (i.e., clinico-pathologic response) for neoadjuvant patients. Patients whose PCR assessment is missing or invalid are counted as not achieving PCR. The PCR status will be summarized for both treatment arms, and its relationship with clinical outcome (e.g., PFS, OS, DOR) will be examined.

Guidance and examples for histopathologic evaluation of surgical specimens after interval cytoreduction are provided in the Ovarian Pathology Handbook and in Appendix 13 of the Protocol.

Analysis of PFS2: PFS2 is defined as the time from randomization to second disease progression, or death from any cause, whichever first. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without second objective disease progression. The method of analysis of PFS2 is analogue to that for the primary endpoint, PFS.

Analysis of Overall Survival at Three Years: The OS rate at three years after randomization will be estimated with the use of the Kaplan-Meier methodology for each treatment arm, along with 95% CIs that are calculated with use of the standard error derived from Greenwood's formula. The 95% CI for the difference in the OS rate between the two treatment arms will be estimated with use of the normal approximation method.

4.5.4.1 Exploratory Patient-Reported Outcomes Analyses

Patient-Reported Outcomes Disease and/or Treatment-Related Symptoms–EORTC Data: Summary statistics (mean, SD, median, and range) of absolute scores and mean changes from the baseline will be calculated for all disease and/or treatment-related symptom items and subscales of the EORTC QLQ-C30 and QLQ-OV28 at each assessment timepoint for each arm during the administration of the treatment and the survival follow-up period. The mean (and 95% CI) and median of the absolute scores and the changes from the baseline will be reported for interval and continuous variables. Previously-published minimally-important differences will be used to identify meaningful

change from the baseline within each treatment group on the disease and/or treatment-related symptoms scales (Osoba et al. 1998; Cocks et al. 2011).

FACT-G, GP5 Single Item Data: A descriptive analysis of absolute scores and the proportion of patients who selected each response option at each assessment visit by treatment arm will be reported for item GP5 ("I am bothered by side effects of treatment") from the Functional Assessment of Cancer Therapy-General (FACT-G) physical well-being subscale. Item GP5 from Version 4 of the FACT-G questionnaire will be scored according to the FACIT scoring manual (Cella 1997). PRO completion, compliance rates, and reasons for missing data will be summarized at each timepoint by treatment arm and surgery status.

Health Economic Data: Health economic data will be used in pharmacoeconomic analyses only and not presented in CSR.

Psychometric Properties of Single Abdominal Pain and Bloating Items—EORTC QLQ-OV28 Data: Analyses will be conducted in all randomized patients to assess whether the psychometric properties of the single abdominal pain and bloating items from the EORTC QLQ-OV28 are comparable to those of the abdominal/GI scale. These analyses will include test-retest reliability, construct validity, and sensitivity/responsiveness to change. Where possible, analyses will be conducted on the combined primary and neoadjuvant chemotherapy/interval surgery patient sample. However, in general, analyses will be split by group to account for the different assessment schedules applied in each patient group. These analyses will mirror, to the extent possible, exploratory psychometric analyses conducted with archival in-house clinical trial data (AURELIA) and thus provide a point of comparison for verification of the psychometric properties of the single abdominal pain and bloating items.

Test-retest reliability will be assessed separately in the primary surgery and neoadjuvant chemotherapy/interval surgery patients from Day 1, Cycle 1 to the first available follow-up (6 and 9 weeks after baseline, respectively) using intraclass correlation coefficients (ICC). Patients will be identified for inclusion in this analysis based on the stability of their scores on the health status item of the EORTC-QLQ-30.

Two forms of construct validity, convergent/divergent and known-groups will be assessed using patients' Day 1, Cycle 1 data. Analyses will be conducted in both the overall randomized patient sample and separately by group. Convergent/divergent validity will be assessed using correlations between the abdominal/GI scale/items and relevant PROs (i.e., EORTC QLQ-C30, EQ-5D, and other scales in the EORTC OV28 module). Known-groups validity will be assessed using t-tests/ANOVA (or non-parametric equivalents, as appropriate) on severity groups defined by scores on the physical, role, and social functioning scales of the EORTC QLQ-C30.

Sensitivity (i.e., responsiveness) to change will be assessed from baseline to multiple follow-up cycles (up to 1 year of survival after treatment discontinuation) using ANOVA (or non-parametric equivalent, as appropriate). Defined change groups will be created using relevant PRO scales. In addition, longitudinal analysis (e.g., multilevel models, generalized estimating equation models) will be conducted to examine change in the abdominal/GI scale/items over the course of the study up to one year of survival after treatment discontinuation. Data will be presented graphically using spaghetti plots to chart each patients' individual growth trajectory.

4.5.4.2 Exploratory Biomarker Analyses

Exploratory biomarker analyses in tumor tissues and plasma, whole blood, or serum will be conducted in an effort to understand the association of these markers with the study drug response, including efficacy and/or AEs. Results will not be presented in the CSR.

4.6 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Maximum and minimum atezolizumab serum concentration data (C_{min} and C_{max}) will be tabulated and summarized. Descriptive statistics will include means, medians, ranges, standard deviations (SDs), and percentage of coefficient of variation (%CV), as appropriate. Descriptive statistics of atezolizumab concentrations by treatment-emergent anti-drug antibody (ADA) status will also be reported.

Additional PK analyses may be conducted if deemed appropriate based on the availability of data. These additional analyses may not be included in the CSR for this study.

4.7 SAFETY ANALYSES

All safety analyses will be performed on the safety population, i.e., all patients who receive any dose of study medication (see Section 4.1).

Safety will be assessed through summaries of AEs, changes in laboratory test results, changes in vital signs, study treatment exposures, and immunogenicity as measured by ADAs and will be presented by treatment arm. Non-overlapping visit windows will be assigned to post-baseline assessments, e.g., assessments falling between Days 22 and 32 will be reported under Week 4.

4.7.1 Exposure of Study Medication

Study treatment (atezolizumab, placebo, paclitaxel, carboplatin, and bevacizumab) exposure, including but not limited to treatment duration, dose intensity, number of cycles and total cumulative dose will be summarized with descriptive statistics as appropriate. The number of missed doses will also be displayed.

4.7.2 Adverse Events

Verbatim description of AEs will be mapped to Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0 (NCI CTCAE v4.0).

Adverse events, occurring on or after the first dose of study treatment will be summarized by mapped term, appropriate thesaurus level, and NCI CTCAE grade regardless of relationship to study drug, as assessed by the investigator. For each patient, if multiple incidences of the same AEs occur, the maximum severity reported will be used in the summaries.

After informed consent has been obtained but prior to initiation of study drug, only SAEs caused by a protocol-mandated intervention should be reported, e.g., SAEs related to invasive procedures such as biopsies.

After the initiation of the study treatment, all AEs will be reported until 30 days after the last dose of the study treatment (which includes study-related surgery) is administered or until the initiation of a new anti-cancer therapy, whichever occurs first. Serious adverse events and AEs of special interest (AESIs) will continue to be reported until 90 days after the last dose of the study treatment (which includes study related surgery) or until starting a new anti-cancer therapy, whichever occurs first. After this reporting period, SAEs believed to be related to prior exposure to study treatment, including study-related surgery, should be reported.

Summary tables including but not limited to the following will be provided:

- AEs
- SAEs
- AEs leading to study treatment discontinuation
- AEs leading to dose reduction or interruption
- Treatment-related AEs
- Severe adverse events (Grade $\geq$3)
- AEs leading to death
- AEs by highest NCI CTCAE Grade
- Sponsor-defined AESI

Multiple occurrences of the same event will be counted once at the maximum severity. All listings of AEs will include all AEs with onset on or after the first study drug treatment up to the data cutoff date.

All deaths and causes of death will be summarized by treatment as well.

4.7.3 Laboratory Data

Laboratory data will be classified according to NCI CTCAE v4.0 and will be summarized descriptively over time including change from baseline. Highest NCI CTCAE grade post-baseline will also be reported and shift tables from baseline to worst value during the study post-baseline will be presented. Of note, abnormal laboratory data that are clinically significant will be reported as AEs and summarized in the AE tables.

A Hy's law analysis will be provided. The potential Hy's law quadrant is defined as the occurrence of either of the following:

- Treatment-emergent ALT or AST $>3\times$ upper limit of normal (ULN) (or baseline value if baseline value was above the ULN) in combination with total bilirubin $>2\times$ ULN (of which $\geq$35% is direct bilirubin).
- Treatment-emergent ALT or AST >3 $\times$ ULN (or baseline value if baseline value was above the ULN) in combination with clinical jaundice.

4.7.4 Vital Signs and ECOG Performance Status

Vital signs will be summarized descriptively by treatment arm over time including change from baseline. ECOG performance status will also be summarized over time.

4.7.5 Anti-Drug Antibody

Incidence of ADAs against atezolizumab will be presented by treatment arm in both the neoadjuvant and primary surgery groups. The analyses of PK and key efficacy and safety data by ADA status will be conducted to explore the potential impact of immunogenicity as appropriate.

4.8 MISSING DATA

The handling of missing data is summarized in Section 4.4 for each endpoint.

4.9 INTERIM ANALYSES

4.9.1 Overall Survival

A total of three analyses of OS will be performed (two interim analyses and one final analysis). The number of events, statistical boundaries (i.e., significance level), and timing of the two interim analyses and the final analysis for OS depending on the results of the definitive analysis of the co-primary endpoint PFS (i.e., final alpha allocated to OS, $0.05\geq\alpha\geq0.046$). The 1st interim analysis will be conducted at the same time as the primary analysis of PFS, which is estimated to be 36 months after first patients enrolled, and when approximately 60% of the total OS events required for the final analysis have occurred. If the pre-specified information fraction of OS analysis is not achieved at the time of the PFS primary analysis, the first interim analysis of OS will still be conducted, and the alpha-spent will be adjusted accordingly. The 2nd interim analysis will occur when approximately 80% of the total OS events required for the final analysis have occurred.

The pre-specified information fraction for OS at first and second interim analyses (i.e., approximately 60% and 80% of the total number of events, respectively) are described in Table 2 below and in Section 4.4.1.

Overall survival will be tested at the significance level determined using the O'Brien-Fleming α-spending function such that the overall type I error rate will be maintained at the 0.05 level (DeMets and Lan 1994).

Table 2 Interim and Final Analyses of Overall Survival

Different Scenarios of PFS Testing	Alpha Level	OS Analysis Type	Time from (months)	Information Fraction (%)	No. of Events	Stopping Boundary (HR)	Stopping Boundary (p-value)
PFS is not statistically significant in either PD-L1+ or ITT	0.046	1st IA	36	60	187 320	≤0.6728 ≤0.7382	≤0.0066
		2nd IA	45	80	249 427	≤0.7447 ≤0.7984	≤0.0200
		FA	56	100	311 534	≤0.7913 ≤0.8364	≤0.0390
PFS is statistically significant in either PD-L1+ or ITT, but not both	0.048	1st IA	36	60	185 317	≤0.6734 ≤0.7393	≤0.0072
		2nd IA	45	80	246 422	≤0.7449 ≤0.7987	≤0.0209
		FA	56	100	308 528	≤0.7920 ≤0.8368	≤0.0407
PFS is statistically significant in both PD-L1+ and ITT	0.050	1st IA	35	60	182 313	≤0.6729 ≤7395	≤0.0075
		2nd IA	44	80	243 418	≤0.7455 ≤0.7995	≤0.0220
		FA	55	100	304 522	≤0.7922 ≤0.8371	≤0.0423

FA=final analysis, FP=first patient in; HR=hazard ratio; IA=interim analysis, ITT=intent-to-treat; OS=overall survival; PD-L1=programmed death-ligand 1; PFS=progression-free survival.

5. CHINA SUBGROUP ANALYSIS

China subgroup includes patients enrolled in global enrollment phase and the China extension phase. The timing of China subgroup analysis depends on the data maturity and pre-specified number of PFS events from China subgroup.

The analysis populations will be equally defined as per Section 4.1, but will only be based on patients who are residents of the People's Republic of China. Analyses of study conduct will be performed as described in Section 4.2. Summaries of demographics, disease history, baseline disease characteristics, and patient treatment history will be produced as described in Section 4.3.

Data from patients who are lost to follow-up or without disease progression or death as of the clinical data cutoff date for China subgroup analysis will be censored at the time of the last tumor assessment when the patient was known to be progression-free (i.e., an overall response other than "progressive disease" or "not evaluable") or at the time of randomization plus one day if there is no post-baseline tumor assessment or all post-baseline tumor assessments have overall responses of "not evaluable." The unstratified Cox proportional hazards model will be used to estimate investigator-assessed PFS HR between the two treatment groups and its 95% CI. An unstratified two-sided log-rank test will provide a comparison for investigator-assessed PFS in this subgroup in a descriptive manner. Plots of the Kaplan-Meier estimate for investigator-assessed PFS will be produced, including medians and confidence limits. The analysis of OS data will be analogous to the ones described above for PFS.

The analysis of other endpoints will use the same methods for global cohorts but with exploratory purpose only.

Safety data for the China subgroup will be analyzed as described in Section 4.7.

6. REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29–41.

Brotto L, Brundage M, Hoskins P, et al. Randomized study of sequential cisplatin-topotecan/carboplatin-paclitaxel versus carboplatin-paclitaxel: effects on quality of life. Support Care Cancer 2016;24:124149. doi: 10.1007/s00520-015-2873-8. Epub: 25 August 2015.

Brundage M, Osoba D, Bezjak A, et al. Lessons learned in the assessment of health-related quality of life: selected examples from the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 2007;25:5078–81.

Cella D. Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System. Version 4. Evanston: Center on Outcomes, Research and Education (CORE), Evanston Northwestern Healthcare and Northwestern University, 1997.

Cocks K. King MT, Velikova G, et al. Evidence-based guidelines for determination of sample size and interpretation of the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol 2011;29:89–96.

DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med 1994;13:1341- 52; discussion 1353-6.

Dmitrienko A, D'Agostino RB Sr, Huque MF. Key multiplicity issues in clinical drug development. Stats Med; 2013;32(7):1079-111.

Fagotti A, Ferrandina G, Vizielli G, et al. Phase III randomised clinical trial comparing primary surgery versus neoadjuvant chemotherapy in advanced epithelial ovarian cancer with high tumour load (SCORPION trial): Final analysis of peri-operative outcome. Eur J Cancer 2016;59:22–33.

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

Fayers PM. Interpreting quality of life data: population-based reference data for theEORTC QLQ-C30. Eur J Cancer 2001a;37:1331–4.

Luckett T, King M, Butow P, et al. Assessing health-related quality of life in gynecologic oncology: a systematic review of questionnaires and their ability to detect clinically important differences and change. Int J Gynecol Cancer 2010;20:664–84. doi: 10.1111/IGC.0b013e3181dad379.

National Cancer Institute. SEER cancer statistics review 1975-2013 [resource on the internet]. 2016. Available from: http://seer.cancer.gov/csr/1975_2013/results_merged/sect_21_ovary.pdf.

Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. Eur J Cancer 2005;41:280–7.

Osoba D, Rodrigues G, Myles J, et al. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16:139–44.

Osoba D. A taxonomy of the uses of health-related quality-of-life instruments in cancer care and the clinical meaningfulness of the results. Med Care 2002;40(6 Suppl):III-31.

Richter R, Oskay-Oezcelik G, Chekerov R, et al. Health-related quality of life during sequential chemotherapy with carboplatin followed by weekly paclitaxel in advanced ovarian cancer: a multicenter phase II study of the North Eastern German Society of Gynecological Oncology. Anticancer Research 2012;32:396–76.

Appendix 1
Protocol Synopsis

TITLE:	**A PHASE III, MULTICENTER, RANDOMIZED STUDY OF ATEZOLIZUMAB VERSUS PLACEBO ADMINISTERED IN COMBINATION WITH PACLITAXEL, CARBOPLATIN, AND BEVACIZUMAB FOR PATIENTS WITH NEWLY-DIAGNOSED STAGE III OR STAGE IV OVARIAN, FALLOPIAN TUBE, OR PRIMARY PERITONEAL CANCER**
PROTOCOL NUMBER:	Roche YO39523 GOG–3015 ENGOT–ov39
VERSION NUMBER:	8
EUDRACT NUMBER:	2016-003472-52
IND NUMBER:	130,637
NCT NUMBER	03038100
TEST PRODUCT:	Atezolizumab (RO5541267)
PHASE:	Phase III
INDICATION:	Ovarian cancer, fallopian tube cancer, primary peritoneal cancer, and cancers of extra-uterine Müllerian origin
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives and Endpoints

This study will evaluate the efficacy and safety of atezolizumab administered with paclitaxel+carboplatin+bevacizumab compared with placebo+paclitaxel+carboplatin+bevacizumab in patients with newly diagnosed, untreated ovarian, fallopian tube, and/or primary peritoneal cancer. Specific objectives and corresponding endpoints for the study are outlined below.

Table 1 Objectives and Corresponding Endpoints

Primary Efficacy Objective	Corresponding Endpoints
• To evaluate the efficacy of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab among all patients and in those patients with PD-L1–positive tumors	• Investigator-assessed PFS, defined as the time from randomization to the occurrence of disease progression, as determined by the investigator from tumor assessments per RECIST v1.1, or death from any cause during the study, whichever occurs first • OS, defined as the time from randomization to death from any cause

Secondary Efficacy Objectives	Corresponding Endpoints
Among patients with measurable residual disease in the primary surgery group:	
• To evaluate the efficacy of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab	• OR, defined as either a CR or PR as determined by the investigator with the use of RECIST v1.1 for patients with measurable residual disease after primary surgery
• To evaluate the duration of efficacy observed with atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab	• DOR, defined as the time interval from first occurrence of a CR or PR to the time of disease progression, as determined by the investigator with the use of RECIST v1.1, or death from any cause, whichever comes first for patients with measurable residual disease after primary surgery
• *To evaluate PROs of function and HRQoL associated with atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab, as measured by the functional andGHS/QoL scales of the EORTC QLQ-C30*	• *Clinically-meaningful improvement, remaining stable, or deterioration in patient-reported function and HRQoL, defined as a ≥ 10-point increase, changes within 10 points, and a ≥ 10-point decrease, respectively, from the baseline score on each of the functional (physical, role, emotional, and social) and GHS/QoL scales of the EORTC QLQ-C30*
Among the patients in the neoadjuvant group:	
• To determine the impact of atezolizumab versus placebo in combination with paclitaxel+carboplatin+bevacizumab on patient-reported abdominal symptoms of OC, as measured by two items from the abdominal/GI symptom scale of the EORTC QLQ-OV28	• Clinically-meaningful improvement in patient-reported abdominal pain or bloating, defined as a ≥ 10-point decrease from the baseline score on either of the two items on the EORTC QLQ-OV28 abdominal/GI symptom scale (items 31 and 32)
• To evaluate PROs of *function and* HRQoL associated with atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab, as measured by the functional and *GHS/*QoL scales of the EORTC QLQ-C30	• Clinically-meaningful improvement in patient-reported function and HRQoL, defined as a ≥ 10-point increase from the baseline score on each of the functional (physical, role, emotional, and social) and *GHS*/QoL scales of the EORTC QLQ-C30
Safety Objective	**Corresponding Endpoints**
• To evaluate the safety and tolerability of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab	• Occurrence and severity of adverse events, with severity determined in accordance to NCI CTCAE v4.0 • Change from baseline in targeted vital signs • Change from baseline in targeted clinical laboratory test results
Pharmacokinetic Objective	**Corresponding Endpoint**
• To characterize the pharmacokinetics of atezolizumab when administered in combination with paclitaxel + carboplatin + bevacizumab	• Minimum and maximum serum concentration of atezolizumab

Table 1 Objectives and Corresponding Endpoints (cont.)

Exploratory Objectives	Corresponding Endpoints
• Among neoadjuvant patients only: To evaluate PCR status and its association with clinical outcomes after administration of atezolizumab versus placebo in combination with paclitaxel + carboplatin + bevacizumab	• Among patients who undergo neoadjuvant therapy prior to interval surgery, PCR status is defined as the clinical amount and histologic characteristics of residual disease assessed at the time of interval cytoreductive surgery
• To evaluate the efficacy of atezolizumab versus placebo administered in combination with paclitaxel + carboplatin + bevacizumab	• The OS rate at 3 years after randomization
• To evaluate PROs of disease and/or treatment-related symptoms associated with atezolizumab versus placebo administered in combination with paclitaxel + carboplatin + bevacizumab, as measured by the EORTC QLQ-C30 and QLQ-OV28	• Mean and mean changes from the baseline score in disease and/or treatment-related symptoms by cycle and between treatment arms as assessed by all symptom items and/or scales of the EORTC QLQ-C30 and QLQ-OV28
• To evaluate any treatment burden patients may experience in association with the addition of atezolizumab to paclitaxel + carboplatin + bevacizumab compared with placebo + paclitaxel + carboplatin + bevacizumab, as measured by a single item (from GP5: "I am bothered by side effects of treatment") from the physical wellbeing subscale of the FACT-G Quality of Life instrument	• Proportion of patients reporting each response option at each assessment timepoint by treatment arm for item GP5 from the FACT-G
• To evaluate and compare between treatment arms patients' health utility as measured by the EQ-5D-5L to generate utility scores for use in economic models for reimbursement	• Health utility scores of the EQ-5D-5L
Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate the immune response from patients who were administered atezolizumab	• The incidence of ADAs against atezolizumab during treatment with atezolizumab administered in combination with paclitaxel + carboplatin + bevacizumab relative to the incidence of ADAs at the baseline
Exploratory Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate the potential effects of ADAs	• The relationship between ADA status and pharmacokinetics, safety, and efficacy

Table 1 Objectives and Corresponding Endpoints (cont.)

Exploratory Biomarker Objective	Corresponding Endpoint
• To assess predictive, prognostic, and pharmacodynamic exploratory biomarkers in archival and/or fresh tumor tissue and blood and their association with disease status and/or patient response to study treatment	• Association of tumor immune-mediated or disease type−related exploratory biomarkers (in archival and/or freshly obtained tumor tissues and plasma, whole blood, or serum) with disease status and/or response to administration of atezolizumab+paclitaxel+carboplatin+bevacizumab; biomarkers may include but are not limited to: CD8 as assessed with the use of IHC Breast cancer susceptibility gene (*BRCA*) status, homologous recombination deficiency, and microsatellite instability as assessed with the use of DNA NGS • Molecular subtyping of ovarian cancer, as assessed by RNA profile • Association of cell-free tumor DNA with tumor burden and treatment response
• To identify biomarkers that are associated with resistance to atezolizumab administered in combination with carboplatin and/or paclitaxel and/or bevacizumab activity, or can increase the knowledge and understanding of disease biology	• The relationship between biomarkers in blood and tumor tissue between pretreatment and post-progression samples collected at the time of disease progression. These biomarkers may include but are not limited to: • Acquired mutations assessed with the use of DNA NGS Changes in the tumor immune microenvironment and biology as assessed by RNA profile and IHC

ADA = anti-drug antibody; CR = complete response; DOR = duration of response; EORTC = European Organisation for Research and Treatment of Cancer; FACT-G = Functional Assessment of Cancer Therapy-General; GI = gastrointestinal; $GHS = global\ health\ status;$ HRQoL = health-related quality of life; IHC = immunohistochemistry; PCR = pathologic and clinical response; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events; NGS = next-generation sequencing; OC = ovarian cancer; OR = objective response; OS = overall survival; PFS = progression-free survival; PR = partial response; PRO =patient-report outcome; QLQ-OV28 = Quality of Life Questionnaire Ovarian Cancer Module 28; RECIST v1.1 = Response Evaluation Criteria in Solid Tumors, Version 1.1.

Study Design

Description of Study

This is a Phase III, global, double-blind, two-arm, randomized study designed to evaluate the efficacy and safety of atezolizumab administered with paclitaxel + carboplatin + bevacizumab compared with placebo + paclitaxel + carboplatin + bevacizumab in patients with Stage III or Stage IV ovarian, fallopian tube, or primary peritoneal cancer with macroscopic residual disease postoperatively (i.e., after primary tumor reductive surgery) or who will undergo neoadjuvant therapy followed by interval surgery. Approximately 1300 patients will be randomized. Patients will be randomized in a 1:1 ratio to one of the two treatment arms.

Eligible patients after primary tumor reductive surgery must be randomized within 42 days of primary surgery postoperatively and commence concurrent treatment with paclitaxel (175 mg/m^2), carboplatin (area under the concentration−time curve at 6 [AUC 6]), and atezolizumab or placebo (1200 mg) on Cycle 1, Day 1. All treatments will be administered

intravenously, and each cycle will be 21 days long. Bevacizumab (15 mg/kg) will be added to paclitaxel, carboplatin, and atezolizumab treatment starting with Cycle 2. After 6 cycles of concurrent chemotherapy+ bevacizumab (5 cycles) + atezolizumab or placebo, patients will continue on maintenance bevacizumab + atezolizumab or bevacizumab +placebo treatment for a total (concurrent + maintenance phases) of 22 cycles of atezolizumab or placebo and 21 cycles of bevacizumab. No additional cycles of maintenance anti-cancer therapy are permitted beyond 22 cycles of therapy in the front-line setting. Study treatment will be discontinued at the time of disease progression, unacceptable toxicity, patient or physician decision to discontinue, death, or study termination by the Sponsor.

Eligible patients in the neoadjuvant therapy group will be randomized prior to the initiation of study treatment. Patients in the neoadjuvant group will commence concurrent treatment with paclitaxel (175 mg/m^2), carboplatin (AUC 6), bevacizumab (15 mg/kg), and atezolizumab or placebo (1200 mg) on Cycle 1, Day 1. All treatments will be administered intravenously, and each cycle will be 21 days long. Cycles 3 and 4 will consist of concurrent paclitaxel (175 mg/m2), carboplatin (AUC 6), and atezolizumab or placebo (1200 mg) with bevacizumab omitted peri-operatively. Interval surgery will occur after Cycle 3 as soon as deemed clinically appropriate, but within a maximum of 6 weeks after receiving Cycle 3. Within 6 weeks after interval surgery, patients will resume concurrent chemotherapy+bevacizumab + atezolizumab or placebo for three more cycles (total of 6 cycles), however bevacizumab treatment will resume at Cycle 5 (i.e., bevacizumab is omitted for Cycle 4). Upon completion of concurrent therapy, patients will commence maintenance treatment with bevacizumab + atezolizumab or bevacizumab+placebo for a total of 22 cycles of atezolizumab or placebo, and 20 cycles of bevacizumab. No additional cycles of maintenance anti-cancer therapy are permitted beyond 22 cycles of therapy in the front-line setting.

After the completion of maintenance therapy (i.e., bevacizumab + atezolizumab or placebo), patients will have an end-of-treatment assessment within 30 days of the last dose of study treatment and then enter the post-treatment follow-up period. The expected study treatment duration for an individual patient is approximately 66 weeks for patients randomized after primary surgery and 70 weeks for patients randomized prior to neoadjuvant therapy.

Progression-free survival (PFS) will be ascertained by the investigator in accordance with radiographic criteria from RECIST v1.1 at fixed intervals. All primary imaging data used for tumor assessment will also be collected by the Sponsor to enable a centralized Independent Review Committee audit of a prespecified subset of PFS data. Overall survival (OS) will be determined by the investigator. Tumor assessments will continue until confirmed disease progression or 5 years after the completion of all study treatment, whichever occurs first. Safety and toxicity assessments will be conducted at each treatment cycle administration.

Patients will undergo a mandatory tumor biopsy sample collection, if clinically feasible, at the time of first evidence of radiographic disease progression according to RECIST v1.1. Cytology from ascites, pleural effusion, and fine needle aspiration (FNA) is not adequate. These samples will be retrospectively analyzed to evaluate and/or characterize pseudoprogression caused by immune cells (ICs) from true progression. In addition, tumor tissue biomarkers related to resistance, disease progression, and clinical benefit of atezolizumab will be analyzed.

Option for additional enrollment in China

Study YO39523 will be enrolling patients globally. Should the Study be opened in mainland China, the Sponsor is targeting a total enrollment of approximately 150 patients. A China extension phase may be initiated if patients enrolled in the global enrollment phase is significantly less than the target. Thus, the China subpopulation will include patients enrolled at sites in China during both the global enrollment phase and the China extension phase. Patients from the China extension phase will be randomized in a 1:1 ratio to the two treatment arms. The patients enrolled in the China extension phase will undergo the same schedule of assessments and will receive paclitaxel, carboplatin, bevacizumab, and atezolizumab or placebo as in the global study. Analyses based on the China subpopulation will be reported separately from the global study.

Number of Patients

Approximately 1300 patients with Stage III or Stage IV ovarian, fallopian tube, or primary peritoneal cancer (i.e., cancers of extra-uterine Müllerian origin) with macroscopic residual disease postoperatively (i.e., after primary tumor reductive surgery) or who will undergo

neoadjuvant therapy followed by interval surgery are expected to be randomized in this study. The enrollment of patients in the neoadjuvant setting will be capped at approximately 20%.

Target Population

Inclusion Criteria

Patients must meet the following criteria for study entry:

- Signed Informed Consent Form (ICF)
- Age ≥ 18 years
- Able to comply with the study protocol, in the investigator's judgment
- Receive a histologic diagnosis of EOC, peritoneal primary carcinoma, or fallopian tube cancer fulfilling the following criteria:

  All epithelial tumors of extra-uterine Müllerian origin by histology (FNA, cytology and/or cell block are not sufficient)

  For patients who will undergo primary tumor reductive surgery: International Federation of Gynecological Oncologists (FIGO) Stage III with gross (macroscopic or palpable) residual disease or Stage IV

    Measurable disease on postoperative imaging studies is not required for eligibility

    FIGO stage is assessed following the completion of initial abdominopelvic surgery that provides the appropriate tissue for histologic evaluation and diagnosis and can be used for exploratory biomarker studies.

    Primary tumor reductive surgery must be within 42 days of randomization.

  For patients who will undergo neoadjuvant treatment and interval surgery: Patients who receive neoadjuvant treatment must also plan to undergo interval surgery after Cycle 3. Mandatory biopsy tissue samples (e.g., core needle or surgically obtained; FNA or cell-block from ascites and/or pleural effusion are inadequate) will be used to confirm histologically that the tumor is of extra-uterine Müllerian origin and to perform exploratory biomarker studies.

    Patients who receive neoadjuvant therapy will include those patients who are not deemed surgically resectable to a state of no gross residual disease due to the extent and/or distribution of disease (e.g., unresectable miliary pattern of peritoneal carcinomatosis, significant diaphragmatic disease, significant involvement of the root of the mesentery, diffuse tumor in the omentum up to the greater curvature of the stomach, extensive miliary carcinomatosis at the root of the mesentery, tumor infiltration of the stomach, surface lesions on the liver).

- Eastern Cooperative Oncology Group (ECOG) performance status of 0, 1, or 2
- Life expectancy > 12 weeks
- Adequate hematologic and end-organ function, defined by the following laboratory test results, obtained within 14 days prior to randomization:

  ANC ≥ 1500 cells/μL (without granulocyte colony stimulating factor support)

  Lymphocyte count ≥ 500/μL

  Platelet count ≥ 100,000/μL without transfusion

  Hemoglobin ≥ 9.0 g/dL

    Patients may be transfused to meet this criterion.

  Serum creatinine ≤ 1.5 × institutional upper limit of normal (ULN)

  Serum bilirubin ≤ 1.5 × ULN

    Patients with known Gilbert disease who have serum bilirubin level ≤ 3 × ULN may be enrolled in the study.

  AST, ALT, and alkaline phosphatase (ALP) ≤ 2.5 × ULN, with the following exceptions:

    Patients with documented liver metastases: AST and/or ALT ≤ 5 × ULN

    Patients with documented liver or bone metastases: ALP ≤ 5 × ULN.

  For patients who do not receive therapeutic anticoagulation: INR or aPTT ≤ 1.5 × ULN

- For patients who receive therapeutic anticoagulation: stable anticoagulant regimen
- Negative hepatitis B surface antigen (HBsAg) test at screening
- Negative total hepatitis B core antibody (HBcAb) test at screening, or positive total HBcAb test followed by a negative hepatitis B virus (HBV) DNA test at screening

  The HBV DNA test will be performed only for patients who have a positive total HBcAb test.
- Negative hepatitis C virus (HCV) antibody test at screening, or positive HCV antibody test followed by a negative HCV RNA test at screening

  The HCV RNA test will be performed only for patients who have a positive HCV antibody test.

  A positive HCV RNA test is sufficient to diagnose active HCV infection in the absence of an HCV antibody test.
- For women of childbearing potential: agreement to remain abstinent (refrain from heterosexual intercourse) or use a contraceptive method with a failure rate of $< 1\%$ per year during the treatment period and for at least 5 months after administration of the last dose of atezolizumab and 6 months after the last dose of bevacizumab, paclitaxel, or carboplatin, whichever is later

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

  Examples of contraceptive methods with a failure rate of $< 1\%$ per year include but are not limited to bilateral tubal ligation and/or occlusion, male sterilization, and intrauterine devices.

  In countries with country-specific health authority mandates, contraceptive methods with a failure rate of $< 1\%$ per year include bilateral tubal ligation/occlusion, male partner sterilization, established and proper use of estrogen-progestin combination hormonal contraceptives that inhibit ovulation, and intrauterine devices/systems.

  The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical study and the preferred and usual lifestyle of the patient.

  Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception.
- Willingness and ability to comply with scheduled visits, treatment plans, laboratory tests, and other study procedures, that include the completion of PRO questionnaires
- Availability of a representative formalin-fixed, paraffin-embedded (FFPE) tumor specimen (screening baseline tissue) in paraffin blocks (preferred) or at least 20 unstained slides.

  Patients with fewer than 20 unstained slides available at baseline may be eligible upon discussion with the medical monitor.

  Tumor tissue should be of good quality based on total and viable tumor content and must be evaluated for PD-L1 expression prior to enrollment. Patients whose tumor tissue is not evaluable for PD-L1 expression are not eligible for enrollment in the study.

  If multiple screening baseline tumor specimens are submitted, patients may be eligible if at least one specimen is evaluable for PD-L1. For the purpose of stratification, the PD-L1 score of the patient will be the maximum PD-L1 score among the samples.

    Acceptable samples include tissue obtained from surgery or core needle biopsies (minimum three cores per paraffin block).

    A paraffin block for FFPE tumor specimens is preferred.

    FNA or cell pellets from ascites or pleural effusion are not acceptable.

    Tumor tissue from bone metastases is not evaluable for PD-L1 expression and is not acceptable.
- For patient enrolled in the extended China enrollment phase: residents in Mainland China, residents in Hong Kong and Taiwan of Chinese ancestry and enrolled at sites recognized by China's National Products Administration (NMPA).

Exclusion Criteria

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

- Received a current diagnosis of borderline epithelial ovarian tumor (formerly tumors of low malignant potential)
- Have recurrent invasive epithelial ovarian, fallopian tube, or primary peritoneal cancer that was treated only with surgery (e.g., patients with Stage IA or Stage IB epithelial ovarian or fallopian tube cancers)

  Patients who received a prior diagnosis of a borderline tumor that was surgically resected and who subsequently developed an unrelated, new invasive epithelial ovarian, fallopian tube, or primary peritoneal cancer are eligible for enrollment, if they have never received prior chemotherapy for any ovarian tumor.
- Have non-epithelial ovarian tumors (e.g., germ cell tumors, sex cord stromal tumors)
- Received prior radiotherapy to any portion of the abdominal cavity or pelvis

  Prior focal radiation for localized cancer of the breast, head and neck, or skin is permitted, if it was completed > 5 years prior to initiation of study treatment, and the patient remains free of recurrent or metastatic disease.
- Received prior chemotherapy for any abdominal or pelvic tumor that include NACT for ovarian, primary peritoneal or fallopian tube cancer

  Patients may have received prior chemotherapy for localized breast cancer, if it was completed > 5 years prior to initiation of study treatment, and the patient remains free of recurrent or metastatic disease.
- Received any biological and/or targeted therapy (including but not limited to vaccines, antibodies, tyrosine kinase inhibitors) or hormonal therapy for management and/or treatment of epithelial ovarian or peritoneal primary cancer
- Have synchronous primary endometrial cancer
- Have a prior history of primary endometrial cancer, except for the following:

  A prior diagnosis of endometrial cancer is allowed if all of the following conditions are met:

  Stage IA cancer

  Superficial myometrial invasion, without lymphovascular invasion

  Grade < 3 or not poorly differentiated subtypes, and this includes papillary serous, clear cell or other FIGO Grade 3 lesions
- With the exception of non-melanoma skin cancer and other specific malignancies as noted above, other invasive malignancies within the last 5 years or previous cancer treatment that contraindicates this protocol therapy (e.g., previous chemotherapy treatment for breast cancer completed > 5 years ago is permitted as per above).
- Are pregnant, lactating, or intend to become pregnant during the study
- Have a history of severe allergic, anaphylactic, or other hypersensitivity reactions to chimeric or humanized antibodies or fusion proteins
- Have a known hypersensitivity or allergy to biopharmaceutical agents produced in Chinese hamster ovary cells or any component of the atezolizumab and/or bevacizumab formulations
- Have an active or history of autoimmune disease or immune deficiency that includes but is not limited to myasthenia gravis, myositis, autoimmune hepatitis, systemic lupus erythematosus, rheumatoid arthritis, inflammatory bowel disease, antiphospholipid antibody syndrome, Wegener's granulomatosis, Sjögren's syndrome, Guillain-Barré syndrome, or multiple sclerosis (see protocol for a more comprehensive list of autoimmune diseases).
- Exceptions include the following:

  Patients with a history of autoimmune-related hypothyroidism on a stable dose of thyroid replacement hormone are eligible.

Patients with controlled Type 1 diabetes mellitus on a stable dose of insulin regimen are eligible.

Patients with eczema, psoriasis, lichen simplex chronicus, or vitiligo with dermatologic manifestations only (e.g., patients with psoriatic arthritis would be excluded) are eligible for enrollment in the study provided that they meet all of the following conditions:

Rash must cover less than 10% of body surface area

Disease is well controlled at baseline and requires only low potency topical steroids

No acute exacerbations of underlying condition within the previous 12 months (i.e., does not require psoralen plus ultraviolet A radiation, methotrexate, retinoids, biologic agents, oral calcineurin inhibitors, high-potency or oral steroids)

- Have a history of idiopathic pulmonary fibrosis, organizing pneumonia (e.g., bronchiolitis obliterans), drug-induced pneumonitis, idiopathic pneumonitis, or evidence of active pneumonitis based on a screening chest computed tomography (CT) scan, with the following exceptions

  A history of radiation pneumonitis in the localized radiation field (fibrosis) is permitted. Patient must still meet all other radiotherapy exclusions listed above.
- Have a positive test result for HIV
- Have active tuberculosis
- Have severe infections within 4 weeks prior to initiation of study treatment, including but not limited to hospitalization for complications of infection, bacteremia, or severe pneumonia
- Have received therapeutic oral or IV antibiotic medication within 2 weeks prior to initiation of study treatment, with the following exceptions:

  Patients who receive prophylactic antibiotic medication (e.g., urinary tract infection prophylaxis, prior to dental procedure) are eligible for enrollment.
- Have significant cardiovascular disease, such as New York Heart Association cardiac disease (Class II or greater), myocardial infarction, or cerebrovascular accident within 3 months prior to initiation of study treatment, unstable arrhythmias, or unstable angina

  Patients with known coronary artery disease, congestive heart failure not meeting the above criteria, or left ventricular ejection fraction $< 50\%$ must be on a stable medical regimen that is optimized in the opinion of the treating physician, in consultation with a cardiologist if appropriate.
- Undergo major surgical procedure within 28 days prior to first bevacizumab dose, or anticipation of the need for a major surgical procedure during the course of the study except patients who receive NACT and will need interval surgery. This may include but is not limited to laparotomy. All incisions should be fully healed, as assessed clinically, prior to starting bevacizumab.

  Consult with the Medical Monitor prior to patient entry for any questions related to the classification of surgical procedures.
- Are administered treatment with a live attenuated vaccine within 4 weeks prior to initiation of study treatment, or anticipation of need for such a vaccine during the course of the study or within 5 months after the last dose of atezolizumab
- Current treatment with anti-viral therapy for HBV
- Have prior allogeneic bone marrow transplantation or solid organ transplant
- Have any other diseases, metabolic dysfunction, physical examination finding, or clinical laboratory finding giving reasonable suspicion of a disease or condition that contraindicates the use of an investigational drug or that may affect the interpretation of the results
- Have any approved or investigational anti-cancer therapy, including chemotherapy or hormonal therapy, with the following exceptions:

  Hormone-replacement therapy or oral contraceptives are allowed.
- Current or recent (within 10 days of initiation of study treatment) use of aspirin (> 325 mg/day) or clopidogrel > 75 mg/day)

- Are administered treatment with any other investigational agent or participation in another clinical study with anti-cancer therapeutic intent
- Have prior treatment with CD137 agonists or immune checkpoint blockade therapies, anti–PD-1, anti–PD-L1, or anti-cytotoxic T-lymphocyte-associated protein 4 therapeutic antibodies
- Have treatment with systemic immunostimulatory agents (including but not limited to interferons [IFNs], interleukin [IL]-2) within 4 weeks or 5 half-lives of the drug, whichever is longer, prior to initiation of study treatment
- Have treatment with systemic immunosuppressive medications (including but not limited to prednisone, cyclophosphamide, azathioprine, methotrexate, thalidomide, and anti–tumor necrosis factor [TNF] agents) within 2 weeks prior to initiation of study treatment, with the following exceptions:

  Patients who have received acute, low-dose, systemic immunosuppressant medications *or a one-time pulse dose of systemic immunosuppressant medication (e.g., 48 hours of corticosteroids for a contrast allergy)* are eligible for enrollment in the study.

  The use of corticosteroids for chronic obstructive pulmonary disease and asthma, mineralocorticoids (e.g., fludrocortisone), *or* low-dose corticosteroids for patients with orthostatic hypotension or adrenocortical insufficiency is allowed.
- Have inadequately-controlled hypertension (defined as systolic blood pressure $>$ 150 mmHg and/or diastolic blood pressure $>$ 100 mmHg)

  Anti-hypertensive therapy to achieve these parameters is allowed.
- Have prior history of hypertensive crisis or hypertensive encephalopathy
- Have significant vascular disease (e.g., aortic aneurysm requiring surgical repair or recent peripheral arterial thrombosis) within 6 months prior to initiation of study treatment
- Have a history of Grade $\geq$ 2 hemoptysis ($\geq$ 2.5 mL of bright red blood per episode) within 1 month prior to screening
- Have evidence of active bleeding, bleeding diathesis, coagulopathy, tumor that involves major vessels (in the absence of therapeutic anticoagulation)
- Have a history or evidence upon physical examination of any CNS disease, including primary brain tumor, CNS metastases, seizures not controlled with standard medical therapy, any brain metastases, or history of cerebrovascular accident (stroke), transient ischemic attack or subarachnoid hemorrhage within 6 months of the first date of treatment on this study.
- History of leptomeningeal disease
- History of Grade $\geq$ 4 venous thromboembolism
- Have current use of full-dose oral or parenteral anticoagulants or thrombolytic agents for therapeutic purposes that has not been stable prior to initiation of study treatment, with the following exceptions:

  The use of full-dose oral or parenteral anticoagulants is permitted as long as the INR or aPTT is within therapeutic limits according to the medical standard of the enrolling institution, and the patient has been receiving a stable dose of anticoagulants prior to initiation of study treatment.

  Prophylactic anticoagulation for the patency of venous access devices is allowed, provided the activity of the agent results in an INR $<$ 1.5 $\times$ ULN and aPTT is within normal limits prior to initiation of study treatment.

  Deep venous thrombosis prophylaxis with low-molecular-weight heparin is permitted.
- Have core biopsy or other minor surgical procedures within 7 days prior to the first dose of bevacizumab

The interval of time between placement of a central vascular access device (CVAD; e.g., Port-a-cath) and the first dose of bevacizumab must be no shorter than 2 days with a well-healed incision

If placing a CVAD between bevacizumab doses, placement must occur at least 14 days from the prior (i.e, pre-CVAD placement) bevacizumab dose, and at least 7 days from the following (i.e., post-CVAD placement) bevacizumab dose.

- Have a history of abdominal fistula or gastrointestinal perforation within 6 months prior to initiation of study treatment, with the following exceptions:

  Patients with granulating incisions healing by secondary intention with no evidence of fascial dehiscence or infection are eligible but require weekly wound examinations.
- Have clinical signs of gastrointestinal obstruction that require routine parenteral hydration, parenteral nutrition, or tube feeding
- Have evidence of abdominal free air not explained by paracentesis or recent surgical procedure
- Have serious, non-healing wound, active ulcer, or untreated bone fracture
- Have proteinuria, as demonstrated by urine dipstick or $>$ 1.0 of protein in a urine protein-to-creatinine ratio and/or 24-hour urine collection

  All patients with $\geq 2+$ protein on dipstick urinalysis at baseline must undergo a urine protein-to-creatinine ratio and/or 24-hour urine collection and demonstrate $\leq$ 1.0 of protein.
- Have known sensitivity to any component of bevacizumab
- Have known sensitivity to any component of paclitaxel
- Have Grade $\geq$ 2 peripheral neuropathy as defined by the National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0 (NCI CTCAE v4.0)
- Have known history of severe allergic reactions to platinum-containing compounds
- Have known history of severe hypersensitivity reactions to products that contain Cremophor® EL (e.g., cyclosporine for injection concentrate and teniposide for injection concentrate)

End of Study

The end of the study is expected to be *approximately 55 months after the first patient is enrolled,* when the approximate preplanned numbers of deaths among the PD-L1–positive patients and the intent-to-treat (ITT) population have been observed. *Deaths will be monitored throughout the course of the study. The sponsor has the right to close the study at any time if futility is observed (e.g., based on the predicted probability of success at the subsequent analysis). The study timelines may be updated as needed.*

Length of Study

The length of study is approximately 55 months from enrollment of the first patient to the "end of study" as described above

Investigational Medicinal Products

The investigational medicinal product (IMP) for this study is atezolizumab.

Atezolizumab

Atezolizumab will be supplied by the Sponsor as a sterile liquid in a single-use, 20-mL glass vial. The vial contains approximately 20 mL (1200 mg) of atezolizumab solution.

Placebo

The placebo will be supplied by the Sponsor and will be identical in appearance to atezolizumab and comprise the same excipients but without atezolizumab drug product.

Non-Investigational Medicinal Products

Bevacizumab will be supplied by the Sponsor. Paclitaxel and carboplatin will be used in commercially available formulations. Bevacizumab will be considered an IMP by local regulations in some countries.

Statistical Methods

Primary Analysis

The primary analysis populations for efficacy are the ITT population, defined as all patients randomized in the study, and the PD-L1–positive subgroup, defined as the patients in the ITT population whose PD-L1 status is IC1/2/3 at the time of randomization. Patients will be grouped in accordance with the treatment assigned at randomization.

The primary analysis population for safety is the Safety Population, defined as all randomized patients who receive at least one dose of the study medication. Patients will be grouped in accordance with the treatment they actually receive, and all patients who received any dose of atezolizumab will be included in the atezolizumab treatment arm for analysis.

If the China extension phase is initiated, data from this phase will not be included in the primary analysis of the main study. A separate analysis will be performed for the China subgroup, where data from the China extension phase and from the patients in the global enrollment phase from China will be combined and summarized (refer to the protocol and the statistical analysis plan [SAP] for further details). All analyses discussed in this section will be restricted to the patients recruited in the global study only, unless noted.

Determination of Sample Size

Approximately 1300 patients will be randomized in the global study in a 1:1 ratio to the two treatment arms.

There are two co-primary efficacy endpoints: investigator-assessed PFS and OS. The overall Type I error rate will be controlled at a two-sided level of 0.05 to test PFS and OS in the PD-L1–positive subgroup and the ITT population. PFS in both the ITT and PD-L1–positive subgroup will be tested in parallel at the same significance level of 0.002 (two-sided). To test OS in both these specified populations, a hierarchical testing approach will be applied. The alpha allocated to OS will be used first to test OS in the PD-L1–positive subgroup. If the significance is reached, the same alpha as used for the PD-L1–positive subgroup OS testing will be passed to OS in the ITT. Note that the OS testing sequence of the populations may be inverted. The OS test in the PD-L1–positive subgroup is initially assigned with an alpha level of 0.046 and the actual alpha level will be determined by the results of the PFS tests. If the PFS test in either the ITT or PD-L1–positive population reaches significance, its assigned alpha of 0.002 will be additively passed to OS.

- The alpha level for OS will be 0.046 if neither of the PFS tests (in the ITT and PD-L1–positive subgroup) reaches significance.
- The alpha level for OS will be 0.048 if only one PFS test (in either the ITT or the PD-L1–positive subgroup) reaches significance.
- The alpha level for OS will be 0.05 if both the PFS tests (in the ITT and the PD-L1–positive subgroup) reach significance.

The sample size of the study is determined by the number of patient deaths required to demonstrate efficacy in terms of OS in the PD-L1–positive subgroup and the ITT population. To detect an improvement in OS with the use of a log-rank test at a two-sided significance level of 0.046, approximately 311 deaths in the PD-L1–positive subgroup will be required to achieve 81% power with a target hazard ratio (HR) of 0.72, and approximately 534 deaths in the ITT population to achieve 80% power with a target HR of 0.78.

Interim Analyses

There will be no interim analysis for PFS. The primary analysis of PFS will take place when approximately 601 PFS events in the ITT and 347 PFS events in the PD-L1–positive subgroup have occurred (whichever is later), which is expected at approximately 36 months after the first patient is enrolled in the study. This provides 90% power to detect a PFS improvement of HR = 0.7 in the ITT, and 91% power in the PD-L1–positive subgroup with HR = 0.62, at a two-sided significance level of 0.002.

Two interim analyses of OS will be performed on patients who are in the ITT and PD-L1–positive populations. The timing of the two interim analyses and the final analysis for OS depends on the results from the primary analysis of the co-primary endpoint of PFS.

The calculation of the sample size and estimates of the analysis timelines are based on the following assumptions:

- PFS and OS are exponentially distributed.
- The median duration of PFS in the control arm is 18 months.
- The median duration of OS in the control arm is 43 months.
- The prevalence of PD-L1–positive (IC1/2/3) patients is 60%.
- The two interim and final analyses of OS use the Lan-DeMets alpha spending function to approximate the O'Brien-Fleming boundary.
- The dropout rate is 5% over 12 months for PFS and OS.
- The recruitment of 1300 patients will take place over 25 months

Appendix 2
Schedule of Assessments for Patients Who Underwent Primary Surgery

	Screening		Concurrent Treatment						Maintenance Treatment																Treatment Completion/ Early Termination [a]	Post-Treatment Follow-up [a]
			Cycle [a]						Cycle [a]																	
	Day –28 to –1	Day –14 to –1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Within 30 days of the last dose of the study treatment	Q3Mo for 2 yr after the treatment completion visit, then Q6Mo for 3 yr, then annually
Signed informed consent [b]	x																									
Demographic data	x																									
Medical history and baseline conditions	x																									
Complete physical examination [c]	x																									
Limited physical examination [d]			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ECOG PS	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Vital signs [e]	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Adverse events [f]		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Concomitant medications		x [g]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Hematology [h]		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Urinalysis [i]		x																								
Urine protein [i]					x		x		x		x		x		x		x		x		x		x		x	

Appendix 2 Schedule of Assessments for Patients Who Underwent Primary Surgery (cont.)

	Screening		Concurrent Treatment						Maintenance Treatment																Treatment Completion/ Early Termination [a]	Post-Treatment Follow-up [a]
			Cycle [a]						Cycle [a]																	
	Day –28 to –1	Day –14 to –1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Within 30 days of the last dose of the study treatment	Q3Mo for 2 yr after the treatment completion visit, then Q6Mo for 3 yr, then annually
Serum chemistry [j]		x	x	x	x	x	x	x	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x [j]	
TSH, free T3, free T4	x					x				x				x				x				x			x	
HIV, HBV, HCV serology [k]	x																									
Urine pregnancy test (for patients of childbearing potential [l]		x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
PT/INR, aPTT		x																								
12-lead ECG	x																									
Radiologic disease assessment [m,n,o]	x				x			x				x				x				x				x	x	x
Serum CA-125 level [p]	x		x	x	x	x	x	x	x		x		x		x		x		x		x		x		x	x
EORTC QLQ-C30, QLQ-OV28, EQ-5D-5L [q]			x		x		x			x				x				x				x			x	x [q]
FACT-G, GP5 single item [r]					x		x			x				x				x				x			x	x [r]
Samples for PK, ADA, and biomarkers	See Appendix 3 for Schedule of PK, ADA, and biomarker sampling																									
Optional whole blood sample for RBR [s]			x																							
Atezolizumab/ placebo administration			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		

Appendix 2 Schedule of Assessments for Patients Who Underwent Primary Surgery (cont.)

	Screening		Concurrent Treatment						Maintenance Treatment																Treatment Completion/ Early Termination [a]	Post-Treatment Follow-up [a]
			Cycle [a]						Cycle [a]																	
	Day –28 to –1	Day –14 to –1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Within 30 days of the last dose of the study treatment	Q3Mo for 2 yr after the treatment completion visit, then Q6Mo for 3 yr, then annually
Bevacizumab administration				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Paclitaxel administration			x	x	x	x	x	x																		
Carboplatin administration			x	x	x	x	x	x																		
Screening baseline FFPE tumor tissue block or 20 unstained slides [t]	x																									
Mandatory FFPE tumor tissue specimen at disease progression [u]																									x	

Appendix 2 Schedule of Assessments for Patients Who Underwent Primary Surgery (cont.)

ADA = anti-drug antibody; CA = cancer antigen; CR = complete response; CT = computed tomography; ECOG PS = Eastern Cooperative Oncology Group performance status; eCRF = electronic Case Report Form; EORTC = European Organisation for Research and Treatment of Cancer; FACT-G = Functional Assessment of Cancer Therapy-General; FFPE = formalin-fixed, paraffin-embedded; HBV = hepatitis B virus; HBcAb = hepatitis B core antibody; HBsAg = hepatitis B surface antigen; HCV = hepatitis C virus; MRI = magnetic resonance imaging; PK = pharmacokinetic; PR = partial response; PRO = patient-reported outcome; Q3Mo = every 3 months; Q6Mo = every 6 months; QLQ-OV28 = Quality of Life Questionnaire Ovarian Cancer Module 28; RECIST = Response Evaluation Criteria in Solid Tumors; UPCR = urine protein-to-creatinine ratio; yr = year.

Note: Study assessments/cycles may be adjusted slightly to accommodate holidays, vacations, and unforeseen major life events. Documentation to justify this decision should be provided.

[a] Each cycle is 21 days. Study drug administration occurs on Day 1 (± 3 days) of each cycle. All other events and assessments during the study treatment period (Cycle 1 through Cycle 22) must occur within 3 days prior to the administration (e.g., activities/assessments for Cycle 1, except the actual drug infusion, must be performed within 3 days before infusion of Cycle 1). The end of the study treatment or early discontinuation visit should occur within 30 days after the last dose of the study treatment is administered. The post-treatment follow-up visits will occur every 3 months (±21 days) for the first 2 years after the treatment completion visit, then every 6 months (±21 days) for 3 years, and then annually (±21 days). Post-treatment visits can be performed via telephone or clinic visit as indicated by assessment requirements.

[b] The Informed consent must be obtained prior to any study-specific procedure and within 28 days (± 7 days) before randomization.

[c] A complete physical examination should include an evaluation of the patient's head, eyes, ears, nose, throat, and the cardiovascular, dermatological, musculoskeletal, respiratory, gastrointestinal, genitourinary, and neurological systems. Any abnormality identified at the baseline should be recorded on the General Medical History and Baseline Conditions eCRF.

[d] Limited, symptom-directed physical examinations should be performed at specified visits and as clinically indicated. Changes from baseline abnormalities should be recorded in the patient notes. New or deteriorated clinically-significant abnormalities should be recorded as adverse events on the Adverse Event eCRF.

[e] Vital signs will include measurements of the respiratory rate, pulse rate, and systolic and diastolic blood pressure while the patient is in a seated position, and body temperature. Vital signs should be measured prior to the first study treatment infusion of the cycle, as clinically indicated, and at other specified timepoints as outlined in the schedule of activities.

Appendix 2 Schedule of Assessments for Patients Who Underwent Primary Surgery (cont.)

[f] All serious adverse events and adverse events of special interest, regardless of their relationship to the study drug, will be reported until 90 days after the last dose of the study drug is administered or the initiation of new anti-cancer therapy, whichever occurs first. All other adverse events, regardless of the relationship to the study drug, will be reported until 30 days after the last dose of the study drug is administered or the initiation of new anti-cancer therapy, whichever occurs first. After this period, the investigator should report any serious adverse events that are believed to be related to prior study drug treatment.

[g] Concomitant medications need to be collected within 7 days prior to starting study treatment.

[h] Hematology consists of CBC, including hemoglobin, WBC count with differential (neutrophils, lymphocytes, eosinophils, monocytes, basophils, and other cells), and platelet count.

[i] Patients must demonstrate negativity for proteinuria (see Section 4.1.2). Urine protein is no longer required for patients who discontinue bevacizumab for reasons other than proteinuria. However, for those who discontinue bevacizumab due to proteinuria, urine protein is required until proteinuria returns to baseline or until the end of the study, whichever occurs first.

[j] Serum chemistry includes BUN, creatinine, sodium, potassium, magnesium, calcium, glucose, total bilirubin, ALT, AST, ALP, and albumin. Serum chemistry eligibility must be confirmed during the screening period before randomization. After Cycle 1, serum magnesium may be performed as clinically indicated. Serum chemistry samples performed during the maintenance phase and at treatment discontinuation may be performed at each cycle, instead of every other cycle, in accordance to country-specific health authority mandates (e.g., Spain).

[k] All patients will be tested for HIV infection prior to enrollment into the study, and HIV-positive patients will be excluded from the study. All patients will be tested for HBsAg and total HBcAb at the baseline. If a patient has a negative HBsAg test and a positive total HBcAb test at screening, an HBV DNA test must also be performed to determine if the patient has an HBV infection. If a patient has a positive HCV antibody test at screening, an HCV RNA test must also be performed to determine if the patient has an HCV infection.

[l] Urine pregnancy tests beyond Cycle 1 for Cycles 2–22, at treatment discontinuation, and at 3 and 6 months after treatment discontinuation must be performed in accordance to country-specific health authority mandates (e.g., Spain, Norway, Czech Republic, Poland, Germany, and Belgium).

[m] All radiographic assessments will be performed according to the time interval (e.g., 9 weeks, 12 weeks, etc.) with the corresponding windows per Section 4.5.5 regardless of the treatment cycle. For patients who have had primary tumor reductive surgery, an initial CT scan or MRI of at least the chest, abdomen, and pelvis is required to establish postsurgical (i.e., primary cytoreductive surgery) baseline of the extent of residual disease within 28 days prior to randomization. In the absence of disease progression, imaging that uses the same modality and encompasses the same field as in the initial pre-treatment evaluation should be repeated (1) every 9 weeks during the concurrent treatment phase, (2) every 12 weeks in the maintenance phase (the first scan in the maintenance phase is performed 12 weeks after the last scan in the concurrent treatment phase), (3) every 3 months for the first 2 years after completion of all protocol therapy, (4) and then every 6 months for the next 3 years, regardless of whether the patient showed measurable disease on the initial CT or MRI. CT scans or MRIs after 5 years of survival follow-up may be done, as clinically indicated. Head/neck imaging at screening is dictated by clinical judgment. If a tumor assessment is performed within 30 days prior to the scan scheduled at the treatment completion/early termination visit, then a specific treatment completion/early termination tumor assessment does not need to be performed.

Appendix 2 Schedule of Assessments for Patients Who Underwent Primary Surgery (cont.)

[n] A CT scan or MRI of at least the patient's chest, abdomen, and pelvis is required to establish the postsurgical (i.e., primary cytoreductive surgery) baseline for the extent of residual disease within 28 days prior to the initiation of study treatment (Scan 1).

[o] Follow-Up Radiographic Assessment of Disease: In the absence of disease progression and/or allergic reaction to IV contrast, imaging with the same modality and field as the initial pretreatment evaluation should be repeated at the following time intervals (i.e., 9 weeks, 12 weeks, etc.) with the corresponding windows per Section 4.5.5 regardless whether the patient had measurable disease on the initial CT or MRI and regardless of the treatment cycle.

Scans 2 and 3: Every 9 weeks from randomization during the concurrent treatment phase (approximately after Cycle 3 and Cycle 6, but may vary).

Scans 4, 5, 6, and 7: Every 12 weeks after the last scan during the concurrent treatment phase, which is now during the maintenance phase (approximately after Cycle 10, Cycle 14, Cycle 18, and Cycle 22).

After the treatment completion visit, every 3 months for 2 years, then every 6 months for 3 years. CT scans or MRIs after 5 years of survival follow-up may be done, as clinically indicated.

[p] Baseline prechemotherapy value is required. When available, also include the presurgical value. CA-125 after 5 years may be done as clinically indicated.

[q] All PRO questionnaires must be completed in their entirety by the patient at the investigational site at the start of the clinic visit before discussion of the patient's health state, laboratory test results, or health records, before the administration of the study treatment, and/or prior to the performance of any other study assessments (e.g., scans) that could bias the patient's responses.

The EORTC QLQ-C30, QLQ-OV28, and EQ-5D-5L questionnaires must be administered and completed by patients in that order at the following assessment timepoints during each treatment and post-treatment period:

Concurrent treatment: At baseline (Cycle 1, Day 1) (±3 days), and on Day 1 (±3 days) of every other cycle thereafter until Cycle 6 (i.e., every 6 weeks)

Maintenance treatment: At Cycle 8, Day 1 (±3 days) and on Day 1 (±3 days) of each cycle every 12 weeks thereafter until Cycle 22

Treatment discontinuation or completion: At the end of the treatment or discontinuation visit within 30 days of the administration of the last dose of the study treatment

Post-treatment follow-up: Patients who discontinue study treatment for progressive disease or loss of clinical benefit or any other reason, regardless of receipt of subsequent anti-cancer therapy will complete the PROs after the treatment completion visit, every 3 months (±21 days) for the first year of the survival follow-up period; every 6 months (±21 days) for the second year of the survival follow-up period; and every year (±21 days) for the final 3 years of the survival follow-up period. Patients in the survival follow-up phase who are unable to visit the site in-person due to COVID-19 may complete the PRO questionnaires via phone interview by site staff.

[r] The single-item GP5 from the FACT-G questionnaire will be the final measure to be administered and must be completed by patients at the following assessment timepoints during each treatment and post-treatment period:

Concurrent treatment: At Cycle 3, Day 1 (±3 days), and on Cycle 5, Day 1 (±3 days)

Maintenance treatment: At Cycle 8, Day 1 (±3 days), and on Day 1 (±3 days) of each cycle every 12 weeks thereafter until Cycle 22

Appendix 2 Schedule of Assessments for Patients Who Underwent Primary Surgery (cont.)

Treatment discontinuation or completion: At the end of the treatment or discontinuation visit within 30 days of the administration of the last dose of the study treatment

Post-treatment follow-up: Patients who discontinue study treatment for progressive disease or loss of clinical benefit or any other reason, regardless of receipt of subsequent anti-cancer therapy will complete FACT-G GP5 after the treatment completion visit, every 3 months (±21 days) for the first year of the survival follow-up period; every 6 months (±21 days) for the second year of the survival follow-up period; every year (±21 days) for the final 3 years of the survival follow-up period. Patients in the survival follow-up phase who are unable to visit the site in-person due to COVID-19 may complete the PRO questionnaires via phone interview by site staff.

[s] Whole blood for DNA isolation will be collected from patients who have consented to optional RBR sampling at Cycle 1, Day 1. If the RBR genetic blood sample is not collected during the scheduled visit, it may be collected after randomization during the clinical study.

[t] Tumor tissue should be of good quality based on the total and viable tumor content (sites will be informed if the quality of the submitted specimen is inadequate to determine tumor PD-L1 status). An FFPE block or at least 20 unstained slides should be provided. Fine-needle aspiration, brushing, cell pellets from pleural effusion, lavage samples, and bone metastases are not acceptable. For core needle biopsy specimens, at least three cores (18 gauge minimum) should be submitted for evaluation. Submission of the screening tumor sample can occur outside the 28-day screening period in conjunction with obtaining the informed consent. Primary surgery must be within 42 days of randomization.

[u] The collection of a sample at disease progression is mandatory if clinically feasible. Preferably, samples collected at the time of radiographic progression should be collected from growing lesions. An FFPE block or at least 15 unstained slides should be provided. Fine-needle aspiration, brushing, cell pellets from pleural effusion, lavage samples, and bone metastases are not acceptable. For core needle biopsy specimens, at no fewer than two cores (18 gauge minimum) should be submitted for evaluation.

Appendix 3
Schedule of Assessments for Patients Who Underwent Neoadjuvant Chemotherapy and Interval Surgery

	Screening		Concurrent Neoadjuvant Treatment			Presurgical/ Surgery Visit	Concurrent Post-neoadjuvant Treatment			Maintenance Treatment																Completion of Treatment/ Early Termination Visit [a]	Post-Treatment Follow-Up [a]
			Cycle [a]				Cycle [a]			Cycle [a]																	
	Day –28 to -1	Day –14 to –1	1	2	3		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Within 30 days of last dose of study treatment	Q3Mo for 2 yr after the treatment completion visit, then Q6Mo for 3 yr, then annually
Written informed consent [b]	x																										
Demographic data	x																										
Medical history and baseline conditions	x																										
Complete physical examination [c]	x																										
Paclitaxel administration			x	x	x		x	x	x																		
Limited physical examination [d]			x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ECOG PS	x		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Vital signs [e]	x		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Adverse events [f]		x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Appendix 3
Schedule of Assessments for Patients Who Underwent Neoadjuvant Chemotherapy and Interval Surgery (cont.)

	Screening		Concurrent Neoadjuvant Treatment			Presurgical/ Surgery Visit	Concurrent Post-neoadjuvant Treatment			Maintenance Treatment																Completion of Treatment/ Early Termination Visit [a]	Post-Treatment Follow-Up [a]
			Cycle [a]				Cycle [a]			Cycle [a]																	
	Day –28 to -1	Day –14 to –1	1	2	3		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Within 30 days of last dose of study treatment	Q3Mo for 2 yr after the treatment completion visit, then Q6Mo for 3 yr, then annually
Concomitant medications [g]		x [f]	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Hematology [h]		x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Urinalysis [i]		x																									
Urine protein [i]				x				x		x		x		x		x		x		x		x		x		x	
Serum chemistry [j]		x	x	x	x		x	x	x	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x	x [j]	x [j]	
TSH, free T3, and free T4	x						x				x				x				x				x			x	
HIV, HBV, HCV serology [k]	x																										
Urine pregnancy test (if childbearing potential exists) [l]		x		x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
PT/INR, aPTT		x																									
12-lead ECG	x																										
Radiographic disease assessment [m,n]					x		x		x				x				x				x				x	x	x
Serum CA-125 level [o]	x		x	x	x		x	x	x	x		x		x		x		x		x		x		x		x	x

Appendix 3
Schedule of Assessments for Patients Who Underwent Neoadjuvant Chemotherapy and Interval Surgery (cont.)

	Screening		Concurrent Neoadjuvant Treatment			Presurgical/ Surgery Visit	Concurrent Post-neoadjuvant Treatment			Maintenance Treatment																Completion of Treatment/ Early Termination Visit [a]	Post-Treatment Follow-Up [a]
			Cycle [a]				Cycle [a]			Cycle [a]																	
	Day –28 to -1	Day –14 to –1	1	2	3		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Within 30 days of last dose of study treatment	Q3Mo for 2 yr after the treatment completion visit, then Q6Mo for 3 yr, then annually
EORTC QLQ-C30, QLQ-OV28, EQ-5D-5L [p]			x			x	x		x		x				x				x				x			x	x [p]
FACT-G, GP5 single item [q]						x			x		x				x				x				x			x	x [q]
Blood samples for PK, ADA , and biomarkers	See Appendix 3 for schedule of PK, ADA, and biomarker sampling.																										
Optional whole blood sample for RBR			x [r]																								
Atezolizumab/ placebo administration			x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Bevacizumab administration			x	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Paclitaxel administration			x	x	x		x	x	x																		

Appendix 3
Schedule of Assessments for Patients Who Underwent Neoadjuvant Chemotherapy and Interval Surgery (cont.)

	Screening		Concurrent Neoadjuvant Treatment			Presurgical/ Surgery Visit	Concurrent Post-neoadjuvant Treatment			Maintenance Treatment															Completion of Treatment/ Early Termination Visit [a]	Post-Treatment Follow-Up [a]
Mandatory FFPE tumor tissue specimen at disease progression [u]																									x	
Carboplatin administration			x	x	x		x	x	x																	
Screening baseline FFPE tumor tissue block or 20 unstained slides [s]	x																									
Biopsy [t]			x																							
FFPE tumor tissue sample from surgical material						x																				
Mandatory FFPE tumor tissue specimen at disease progression [u]																									x	

ADA = anti-drug antibody; CA = cancer antigen; CR = complete response; CT = computed tomography; ECOG PS = Eastern Cooperative Oncology Group performance status; eCRF = electronic Case Report Form; EORTC = European Organisation for Research and Treatment of Cancer; FACT-G = Functional Assessment of Cancer Therapy-General; FFPE = formalin-fixed, paraffin-embedded; HBV = hepatitis B virus; HBcAb = hepatitis B core antibody; HBsAg = hepatitis B surface antigen; HCV = hepatitis C virus; MRI = magnetic resonance imaging; PK = pharmacokinetic; PR = partial response; PRO = patient-reported outcome; Q3Mo = every 3 months; Q6Mo = every 6 months;; QLQ-OV28 = Quality of Life Questionnaire Ovarian Cancer Module 28; RECIST = Response Evaluation Criteria in Solid Tumors; UPCR = urine protein-to-creatinine ratio; $yr = year$.

Note: Study assessments/cycles may be adjusted slightly to accommodate holidays, vacations, and unforeseen major life events. Documentation to justify this decision should be provided.

[a] Each cycle is 21 days. Study drug administration occurs on Day 1 (± 3 days) of each cycle. All other events and assessments during the study treatment period (Cycle 1 through Cycle 22) must occur within 3 days prior to the administration (e.g., activities/assessments for Cycle 1, except the actual drug infusion, must be performed within 3 days before infusion of Cycle 1). The end of the study treatment or early discontinuation

visit should occur within 30 days after the last dose of the study treatment is administered. The post-treatment follow-up visits will occur every 3 months (±21 days) for the first 2 years after the treatment completion visit, then every 6 months (±21 days) for 3 years, and then annually (±21 days). Post-treatment visits can be performed via telephone or clinic visit as indicated by assessment requirements.

[b] The Informed consent must be obtained prior to any study-specific procedure and within 28 days (± 7 days) before randomization.

[c] A complete physical examination should include an evaluation of the patient's head, eyes, ears, nose, throat, and the cardiovascular, dermatological, musculoskeletal, respiratory, gastrointestinal, genitourinary, and neurological systems. Any abnormality identified at the baseline should be recorded on the General Medical History and Baseline Conditions eCRF.

[d] Limited, symptom-directed physical examinations should be performed at specified visits and as clinically indicated. Changes from baseline abnormalities should be recorded in the patient notes. New or deteriorated clinically-significant abnormalities should be recorded as adverse events on the Adverse Event eCRF.

[e] Vital signs will include measurements of the respiratory rate, pulse rate, and systolic and diastolic blood pressure while the patient is in a seated position, and body temperature. Vital signs should be measured prior to the first study treatment infusion of the cycle, as clinically indicated, and at other specified timepoints as outlined in the schedule of activities.

[f] All serious adverse events and adverse events of special interest, regardless of their relationship to the study drug, will be reported until 90 days after the last dose of the study drug is administered or the initiation of new anti-cancer therapy, whichever occurs first. All other adverse events, regardless of the relationship to the study drug, will be reported until 30 days after the last dose of the study drug is administered or the initiation of new anti-cancer therapy, whichever occurs first. After this period, the investigator should report any serious adverse events that are believed to be related to prior study drug treatment.

[g] Concomitant medications need to be collected within 7 days prior to starting study treatment.

[h] Hematology consists of CBC, including hemoglobin, WBC count with differential (neutrophils, lymphocytes, eosinophils, monocytes, basophils, and other cells), and platelet count.[i] Patients must demonstrate negativity for proteinuria (see Section 4.12). Urine protein is no longer required for patients who discontinue bevacizumab for reasons other than proteinuria. However, for those who discontinue bevacizumab due to proteinuria, urine protein is required until proteinuria returns to baseline or until the end of the study, whichever occurs first.

[j] Serum chemistry includes BUN, creatinine, sodium, potassium, magnesium, calcium, glucose, total bilirubin, ALT, AST, ALP, and albumin. Serum chemistry eligibility must be confirmed during the screening period before randomization. After Cycle 1, serum magnesium may be performed as clinically indicated. Serum chemistry samples performed during the maintenance phase and at treatment discontinuation may be performed at each cycle, instead of every other cycle, in accordance to country-specific health authority mandates (e.g., Spain).

[k] All patients will be tested for HIV infection prior to enrollment into the study, and HIV-positive patients will be excluded from the study. All patients will be tested for HBsAg and total HBcAb at the baseline. If a patient has a negative HBsAg test and a positive total HBcAb test at screening, an HBV DNA test must also be performed to determine if the patient has an HBV infection. If a patient has a positive HCV antibody test at screening, an HCV RNA test must also be performed to determine if the patient has an HCV infection.

[l] Urine pregnancy tests beyond Cycle 1 for Cycles 2–22, at treatment discontinuation, and at 3 and 6 months after treatment discontinuation must be performed in accordance to country-specific health authority mandates (e.g., Spain, Norway, Czech Republic, Poland, Germany, and Belgium).

[m] All radiographic assessments will be performed according to the time interval (e.g., 9 weeks, 12 weeks, etc.) with the corresponding windows per Section 4.5.5 regardless of the treatment cycle. The initial CT scan/MRI of at least the patient's chest, abdomen, and pelvis is required within

Appendix 3
Schedule of Assessments for Patients Who Underwent Neoadjuvant Chemotherapy and Interval Surgery (cont.)

28 days of randomization to establish the first baseline prior to starting neoadjuvant therapy and interval cytoreductive surgery. A CT scan or MRI of at least the patient's chest, abdomen, and pelvis is required to establish the new baseline (i.e., postsurgical from interval cytoreductive surgery) for the extent of residual disease within 14 days prior to re-initiation of study treatment (Scan 3). CT scans or MRIs after 5 years of survival follow-up may be done as clinically indicated. Head/neck imaging at screening is dictated by clinical judgment. If a tumor assessment is performed within 30 days prior to the scan scheduled at the treatment completion/early termination visit, then a specific treatment completion/early termination tumor assessment does not need to be performed.

[n] Follow-Up Radiographic Assessment of Disease: In the absence of disease progression and/or allergic reaction to IV contrast, imaging with the same modality and field as the initial pretreatment evaluation should be repeated at the following time intervals (i.e., 9 weeks, 12 weeks, etc.) with the corresponding windows per Section 4.5.5 regardless whether the patient had measurable disease on the initial CT or MRI and <u>regardless of the treatment cycle:</u>

Concurrent treatment phase Scan 2: 9 weeks from randomization during the concurrent treatment phase (approximately end of Cycle 3).

Scan 3: postsurgical scan done after interval cytoreductive surgery and as close as possible (within 14 days) to starting Cycle 4 during the concurrent treatment phase.

Scan 4: 9 weeks from Scan 3 (i.e., postsurgical scan done after interval cytoreductive surgery) during the concurrent treatment phase (approximately end of Cycle 6). Tumor assessment scans during the concurrent treatment phase must be done every 9 weeks, regardless of treatment cycle.

Maintenance phase Scans 5, 6, and 7: Every 12 weeks after the last scan during the concurrent treatment phase, which is now during the maintenance phase (approximately after Cycle 10, Cycle 14, Cycle 18, and Cycle 22).

After the treatment completion visit, every 3 months for 2 years, then every 6 months for 3 years. CT scans or MRIs after 5 years of survival follow-up may be done, as clinically indicated.

[o] Baseline prechemotherapy value is required. When available, also include the presurgical value. CA-125 after 5 years may be done as clinically indicated.

[p] All PRO questionnaires must be completed in their entirety by the patient at the investigational site at the start of the clinic visit before discussion of the patient's health state, laboratory test results, or health records, before the administration of the study treatment, and/or prior to the performance of any other study assessments (e.g., scans) that could bias the patient's responses.

The EORTC QLQ-C30, QLQ-OV28, and EQ-5D-5L questionnaires must be administered and completed by patients in that order at the following assessment timepoints during each treatment and post-treatment period:

Concurrent Treatment: At baseline (Cycle 1, Day 1) (± 3 days); post Cycle 3/pre-Interval Surgery visit (± 3 days); post-Interval Surgery at Cycle 4, Day 1 (±3 days); Cycle 6, Day 1 (± 3 days)

Maintenance Treatment: At Cycle 8, Day 1 (± 3 days) and on Day 1 (± 3 days) of each cycle every 12 weeks thereafter until Cycle 22

Treatment Discontinuation or completion: At the end of treatment or discontinuation visit within 30 days of the administration of the last dose of the study treatment

Appendix 3
Schedule of Assessments for Patients Who Underwent Neoadjuvant Chemotherapy and Interval Surgery (cont.)

Post-Treatment Follow-up: Patients who discontinue study treatment for progressive disease or loss of clinical benefit or any other reason, regardless of receipt of subsequent anti-cancer therapy will complete the PROs after the treatment completion visit, every 3 months (±21 days) for the first year of the survival follow-up period; every 6 months (±21 days) for the second year of the survival follow-up period; and every year (±21 days) for the final 3 years of the survival follow-up period. Patients in the survival follow-up phase who are unable to visit the site in-person due to COVID-19 may complete the PRO questionnaires via phone interview by site staff.

[q] The single-item GP5 from the FACT-G questionnaire will be the final measure to be administered and must be completed by patients at the following assessment timepoints during each treatment and post-treatment period:

Concurrent treatment: At the post Cycle 3/pre-Interval Surgery visit (±3 days), and on Cycle 6, Day 1 (±3 days)

Maintenance treatment: At Cycle 8, Day 1 (±3 days), and on Day 1 (±3 days) of each cycle every 12 weeks thereafter until Cycle 22

Treatment discontinuation or completion: At the end of the treatment or discontinuation visit within 30 days of the administration of the last dose of the study treatment

Post-treatment follow-up: Patients who discontinue study treatment for progressive disease or loss of clinical benefit or any other reason, regardless of receipt of subsequent anti-cancer therapy will complete FACT-G GP5 after the treatment completion visit, every 3 months (±21 days) for the first year of the survival follow-up period; every 6 months (±21 days) for the second year of the survival follow-up period; every year (±21 days) for the final 3 years of the survival follow-up period. Patients in the survival follow-up phase who are unable to visit the site in-person due to COVID-19 may complete the PRO questionnaires via phone interview by site staff.

[r] Whole blood for DNA isolation will be collected from patients who have consented to optional RBR sampling at Cycle 1, Day 1. If the RBR genetic blood sample is not collected during the scheduled visit, it may be collected after randomization during the clinical study.

[s] Tumor tissue should be of good quality based on the total and viable tumor content (sites will be informed if the quality of the submitted specimen is inadequate to determine tumor PD-L1 status). An FFPE block or at least 20 unstained slides should be provided. Fine-needle aspiration, brushing, cell pellets from pleural effusion, lavage samples, and bone metastases are not acceptable. For core needle biopsy specimens, at least three cores (18 gauge minimum) should be submitted for evaluation. Submission of the screening tumor sample from needle core biopsy or laparoscopy can occur outside the 28-day screening period in conjunction with obtaining the informed consent.

[t] For patients who are enrolled in the biomarker cohort; tumor biopsy samples should be collected by core needle (18 gauge minimum) or excisional biopsy at Cycle 1, Day 15. Three cores per paraffin block should be submitted for analysis.

[u] The collection of a sample at disease progression is mandatory if clinically feasible. Preferably, samples collected at the time of radiographic progression should be collected from growing lesions. An FFPE block or at least 15 unstained slides should be provided. Fine-needle aspiration, brushing, cell pellets from pleural effusion, lavage samples, and bone metastases are not acceptable. For core needle biopsy specimens, at no fewer than two cores (18 gauge minimum) should be submitted for evaluation.

Appendix 4
Schedule of Pharmacokinetic, Immunogenicity, and Biomarker Samples [a]

		Sample Type	
Visit	Timepoint	Patients who undergo primary surgery	Patients who undergo NACT interval surgery (and, where applicable, the biomarker cohort only; n = 100)
Screening (Day −28 to Day −1)	—	—	—
Randomization	—	Atezolizumab/placebo biomarker (tissue)	Atezolizumab/placebo biomarker (tissue)
Cycle 1, Day 1	Prior to the first infusion (up to 24 hours prior)	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b] Atezolizumab/placebo biomarker (blood, plasma [c])	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b] Atezolizumab/placebo biomarker (blood, plasma [c])
	30 minutes (± 15 minutes) after end of infusion	Atezolizumab/placebo pharmacokinetics (serum) [b]	Atezolizumab/placebo pharmacokinetics (serum) [b]
Cycle 1, Day 15	—	—	Atezolizumab/placebo biomarker (tissue) (biomarker cohort only; n = 100) Atezolizumab/placebo biomarker (plasma [c]) (biomarker cohort only; n = 100)
Cycle 2	Prior to the first infusion (up to 24 hours prior)	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b] Atezolizumab/placebo biomarker (plasma [c])	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b] Atezolizumab/placebo biomarker (plasma [c])
Cycle 3	Prior to the first infusion (up to 24 hours prior)	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]
	30 minutes (± 15 minutes) after the end of infusion	Atezolizumab/placebo pharmacokinetics (serum) [b]	Atezolizumab/placebo pharmacokinetics (serum) [b]

Appendix 4
Schedule of Pharmacokinetic, Immunogenicity, and Biomarker Samples [a] (cont.)

		Sample Type	
Visit	Timepoint	Patients who undergo primary surgery	Patients who undergo NACT interval surgery (and, where applicable, the biomarker cohort only; n = 100)
Cycle 4	Prior to the first infusion (up to 24 hours prior)	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b] Atezolizumab/placebo biomarker (plasma [c, d])	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b] Atezolizumab/placebo biomarker (plasma [c, d])
Cycle 8 and Cycle 16	Prior to the first infusion (up to 24 hours prior)	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]	—
At surgical visit	—	—	Atezolizumab/placebo biomarker (tissue) Atezolizumab/placebo biomarker (plasma [c, d])
At disease progression[e]	—	Atezolizumab/placebo biomarker (tissue, plasma [c, d])	Atezolizumab/placebo biomarker (tissue, plasma [c, d])
Treatment discontinuation/ completion visit [f]	(± 24 hours)	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]	Atezolizumab/placebo biomarker (plasma [c, d]) Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]
≥ 90 days after last dose of atezolizumab[h]	(± 1 week[g])	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]	Atezolizumab/placebo pharmacokinetics (serum) [b] Atezolizumab/placebo ADA (serum) [b]

Appendix 4
Schedule of Pharmacokinetic, Immunogenicity, and Biomarker Samples [a] (cont.)

ADA = anti-drug antibody; NACT = neoadjuvant chemotherapy.

Note: Study assessments/cycles may be adjusted slightly to accommodate holidays, vacations, and unforeseen major life events. Documentation to justify this decision should be provided.

[a] Pharmacokinetic and ADA samples will not be collected for the patients enrolled in the sites in China.

[b] Patients who discontinue atezolizumab/placebo no longer need to provide pharmacokinetics/ADA samples.

[c] EDTA plasma.

[d] This may include WGS and/or RBR samples. Refer to the laboratory manual for details.

[e] Samples should be collected from either RECIST progression or investigators-determined clinical progression

[f] Patients who complete the study treatment period will return to the clinic for a treatment completion visit at 30 days from the administration of the last dose of the study drug. Patients who discontinue the study drug prematurely will return to the clinic for a treatment discontinuation visit 30 days after the administration of the last dose of the study drug. The visit at which the response assessment shows progressive disease may be designated as the treatment discontinuation visit.

[g] Samples greater than one week are acceptable.

[h] Can be combined with the Month 3 Survival Follow-Up visit when appropriate.

Appendix 5 Blinded Independent Central Review Audit Plan

BLINDED INDEPENDENT CENTRAL REVIEW AUDIT PLAN FOR YO39523

TITLE:	**A PHASE III, MULTICENTER, RANDOMIZED, STUDY OF ATEZOLIZUMAB VERSUS PLACEBO ADMINISTERED IN COMBINATION WITH PACLITAXEL, CARBOPLATIN, AND BEVACIZUMAB TO PATIENTS WITH NEWLY-DIAGNOSED STAGE III OR STAGE IV OVARIAN, FALLOPIAN TUBE, OR PRIMARY PERITONEAL CANCER**
PROTOCOL NUMBER:	YO39523 / GOG-3015 / ENGOT-ov39
STUDY DRUG:	Atezolizumab (RO5541267)
VERSION NUMBER:	1
IND NUMBER:	130637
EUDRACT NUMBER:	2016-003472-52
SPONSOR:	F. Hoffmann-La Roche Ltd
PLAN PREPARED BY:	[redacted]
DATE FINAL:	2 March 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

1. PURPOSE OF DOCUMENT

The purpose of this document is to describe the audit plan related to the blinded independent central review (BICR)-determined progression-free survival which will be conducted in support of clinical trial protocol YO39523 [11].

In addition to the audit strategy and methodology, this document also describes the roles and responsibilities of participating parties, and some implementation details. Revisions or amendments to the protocol will not require an amendment to this document unless substantive changes impact the blinded independent central review (BICR).

2. BACKGROUND

2.1 BLINDED INDEPENDENT CENTRAL REVIEW FOR PROGRESSION-FREE SURVIVAL

Progression-free survival is commonly used as a primary endpoint for evaluating efficacy of cancer products in clinical trials pursuing marketing approval. When progression-free survival is the primary efficacy endpoint of a clinical trial, a blinded independent central review is usually required in addition to local-evaluated (LE) progression-free survival.

Local-evaluated progression-free survival could potentially be biased in favor of the IMP treatment especially in open-label studies when investigators know their patients' treatment assignment. Blinded independent central review is implemented to reduce the potential bias from the local assessments. For blinded studies, there is usually less concern on potential bias from LE. Blinded independent central review-assessed progression-free survival is particularly relevant for cancers with diffuse intraperitoneal spread patterns and/or frequent non-target lesions, such as ovarian cancer. Indeed, variation between evaluators' assessments may be particularly germane in ovarian cancer, where diffuse peritoneal implantation and frequent non-target lesions, including ascites, can complicate radiologic assessments.

2.2 AUDIT FOR BLINDED INDEPENDENT CENTRAL REVIEW

The use of blinded independent central review (BICR) increases trial complexity and cost, and it is an added burden to both investigators and sponsors. Subjectivity in disease progression assessments and differences in evaluations between LE and BICR at the individual patient level exist. However, some publications [1] and FDA [2,3] analyses have suggested that there is high correlation between LE and BICR assessments with respect to the relative treatment effect measure (hazard ratio). A

meta-analysis conducted by the Pharmaceutical Research and Manufacturers of America (PhRMA) and FDA showed that some degree of discordance between local and central readers may not affect the conclusions about the efficacy of a study as measured by the HR. This high correlation raises the question whether complete case BICR evaluation is needed.

Dodd (2008) [4] and Amit (2011) [5] suggest using BICR as an auditing tool to assess the reliability of marginally positive results or to use a sample-based BICR in not adequately blinded studies. In their view, a full BICR would only be required for smaller trials or in situations, in which confidence in the local evaluation results needs to be increased. Retrospective analyses of 27 randomized clinical trials in various indications from Zhang (2012) [1] suggest that an audit strategy is also feasible.

An oncologic advisory committee discussion (ODAC) meeting was held in July 2012 to discuss whether the current practice of complete-case BICR should be replaced by a random sample-based BICR audit, based on the information and analyses presented herein [2].In the ODAC, focusing on a general discussion of oncologic products seeking marketing approval in the United States for the treatment of non-hematologic malignancies, members felt that the available data support a sample-based BICR [3]. Committee members agreed that a random sample-based BICR audit should be considered: BICR audit on a random sample of clinical trial patients may streamline the process and reduce the drug development burden. However, the potential merits must be viewed in tandem with the potential limitations and challenges: no definite conclusions were drawn. They also advised against the complete elimination of BICR, which could jeopardize the integrity of the local site evaluation. Concerns were raised that in small sized trials, studies with a moderate PFS improvement, and certain difficult-to-assess tumor types, this sample approach would not be adequate.

Although analyses [Dodd (2008) [4], Amit (2011) [5], and Zhang (2012) [1] have showed that a BICR audit to assess potential bias in the local site evaluation is a feasible approach, the logistics of how the audit should proceed need further discussion and consideration. The method of selecting the random sample audit needs to be determined; a simple random sample audit may not be sufficient, for example, to ensure representation of all study sites. Efforts should also be made to minimize any additional burden that the audit may cause the investigator or sponsor without compromising the integrity of the study. Selection of the actual audit strategy to implement within a trial may need to be determined on a case-by-case basis. At the ODAC meeting held on July 2012, examples of two methods have been proposed to use a random sample-based IRC audit to examine potential systematic bias.

2.3 AUDIT METHODS

Although measurement error is inherent in the reading of radiographic scans, regulatory considerations for drug approval are based on the relative treatment effect at the population level. Given that multiple publications have corroborated the strong correlation between local site evaluation– and BICR-assessed PFS treatment effect estimates, there is a need for the exploration of alternative strategies to detect bias in the local site evaluation. The results of the analyses presented in Zhang (2013) [1] support that a random sample-based BICR audit is a viable alternative to a complete-case BICR and may be a more efficient strategy for bias evaluation of the local site evaluation. In Zhang (2013) [8], the performance characteristics of the two audit methods proposed in the literature are evaluated on 26 prospective, randomized phase III registration trials in non-hematologic malignancies. The results support that a BICR audit to assess potential bias in the local site evaluation is a feasible approach. Note that, although there is general agreement, in one study reported in Zhang (2013), there were divergent conclusions depending on the choice of local site evaluation and BICR, justifying a continued need for BICR to provide assurance about the local site evaluation–based treatment effect.

Originally, two methods have been proposed for this type of audit. [Dodd (2011) [6] and Amit (2011)] Dodd uses the HRs directly and is about early testing of the statistical significance of the BICR based on a sample – so it may, in the case of a large treatment effect, accept the sample even if the HRs estimated are quite different. In Dodd's method, the treatment effect based on the full BICR evaluation can be estimated on the basis of investigator evaluation and the correlation between investigator assessment and BICR assessment of the audit sample. A full BICR evaluation is only needed to confirm the investigator assessed treatment effect is reliable when the upper limit of the confidence interval (CI) of the HR estimate based on the BICR and investigator assessment of the audit sample exceeds a pre-specified clinical irrelevance factor (CIF). The CIF is a factor reflecting the smallest treatment effect that is clinically meaningful. The size of the audit sample required depends on the number of PFS events, the magnitude of the effect, the CIF, and the correlation between the BICR and investigator assessment. More details about calculating the audit sample size can be found in Dodd et al. (2011). Amit et al. (2011) proposed using differential discordance as a measure to detect potential systematic evaluation bias in investigator-assessed tumor data (the PhRMA approach). Two types of discrepancy measure can be calculated for each treatment arm: early discrepancy rate (EDR) is defined as the frequency with which the investigator declares progression early relative to BICR as a proportion of the total number of investigator-assessed progression events; late discrepancy rate (LDR) is defined as the frequency with which the investigator declares progression later than BICR as a proportion of the total number of discrepancies within the arm. Differential discordance for each type of measure can be defined as the LDR/EDR in the experimental arm minus the rate in the control arm. The former was undertaken by the national cancer institute (NCI), while the latter was undertaken by the PhRMA PFS

working group. Both the PhRMA and NCI approaches aim to provide a decision rule as to whether a full BICR is needed by comparing LE and BICR data in the audit sample. However, the philosophy of both approaches does somewhat differ. The NCI approach uses the LE and BICR hazard ratio (HR) directly and tests whether the sample BICR is sufficient to already conclude that the full sample BICR would likely be statistically significant. Whereas, the PhRMA approach focuses on whether there is evidence of bias in the assessment of investigators by a between arm comparison of carefully chosen discordance metrics. These two approaches may come to different recommendations for the same trial or pick out different trials for complete BICR. Challenge of selection of sample size and threshold exists.

Dodd (2011) method seems to perform well in most situations; that is, it seems able to distinguish between trials with and without bias present. The savings with respect to audit size varies from case to case, depending on the study size and magnitude of the observed local site evaluation PFS treatment effect; larger effect sizes and/or studies typically require smaller audit proportions. Method from Amit (2011) is intuitively appealing, but needs further evaluation, particularly with respect to determination of the appropriate threshold value. Potential reasons for its variable performance that need further evaluation include loss of important information due to dichotomization and ignoring patients who were censored by both the local site evaluation and BICR in the definition of EDR and LDR. This method counts discordances but not how far apart they are. For example, with the LDR, the local site evaluation could call progressive disease right after BICR or a long time after. If the late discrepancies occurred in the control arm many visits after BICR, this would produce more significant bias in the HR estimate. On a similar note, we could have a relatively small number of late discrepancies, but if they occur very late, then this would produce greater bias in the HR estimate.

In 2015, to deepen the understanding of the two proposed audit methods, the PhRMA approach (i.e., Amit (2011)) and the NCI approach (i.e., Dodd (2011)), as described above and to collect first-hand operational experience, Roche retrospectively performed PFS audit for study TH3RESA after its primary analysis of PFS by LE.[7] Roche implemented the PhRMA approach (i.e., Amit (2011)) as the primary audit approach and the strategy proposed by Dodd (2011) (i.e., the NCI approach) as a secondary approach for TH3REASA study. It was not clear which audit strategy was more superior and more appropriate for this study, because it was noted that the statistical properties of the audit strategies and the proper selection of audit sample size and threshold require further investigation to be better understood. For the PhRMA approach, the difficulty lies in the challenge of translating differential discordance rate for a specific study between two arms to bias in PFS, which is the real objective we are interested in, and the selection of threshold in terms of discordance rates. For the NCI approach, the selection of the threshold (i.e., CIF) is somewhat subjective.

Later, a third method [4], developed by Andrew Stone, which shares some key features with both the Dodd and Amit methods, was published, see section 5.1 for more details.

The authors labelled their method model-free. Whilst asymptotic assumptions of normality are made there is no requirement, as per the Amit method, for the construction of an underlying, study-specific error model. The method was labelled as model-free as it is an extension of the approach, described as parameter-free by Jennison and Turnbull[14], that can be used in group sequential trials for the construction of futility analyses. In common with the Dodd method, it utilizes the HRs directly, and in common with the Amit method, it aims to detect bias in terms of differences in treatment effect estimates rather than to retest statistical significance. A key advantage of this approach is that the sensitivity and specificity of using the audit sample to detect if there is any assessment bias by LE can be mostly mathematically computed.

2.4 AUDIT METHODS ASSESSMENTS

Roche ran a few simulations based on similar error models described in Dodd (2011) and Amit (2011) to compare the 3 methods described above. Across the simulations with assumptions of various magnitudes of bias between LE and BCIR and various audit samples sizes, it consistently shown that test statistics HRR gave the optimal performance in the sense of specificity and sensitivity based on the ROC curves compared to test statistics DD of LDR did. Details are provided in Appendix 1. Thus, HRR will be used as the test statistics for study YO39523 PFS audit, and the threshold for the HRR test statistics will be determined based on the approach introduced in Stone (2015) and will be described in details in section 5.

Simulation was also ran to assess the performance of two sampling strategies: the simple random sampling and interim sampling. Simple random sampling includes all images of patients who are randomly sampled from the ITT population for audit by BICR review. Interim sampling includes all images of all patients up to an interim time point for audit. The simulation shows comparable performance in the sense of sensitivity and specificity, when the numbers of images included for audit are about the same. More details are included in Appendix 1. One challenge implementing the “simple random sampling” strategy is that the audit can be performed only after all images are available (i.e. after database lock). This may potentially elongate timeline of the primary analysis of PFS, especially if the pre-specified threshold is crossed and a full BICR is needed. Interim sampling can avoid this aforementioned challenge, since the audit can be performed at the interim time points and the audit results will be available at that time as well. If the audit at interim is passed, then BICR of future images won’t be needed any more. If the audit at interim failed to rule out potential bias, the BICR will be continued. Neither case will elongate the post database lock timeline. Thus, interim sampling will be used for study YO39523 PFS audit. Details are described in section 4.

3. CLINICAL TRIAL YO39523

3.1 STUDY DESIGN

Clinical trial YO39523 is a Phase III randomized double-blind, placebo-controlled study evaluating the efficacy and safety of paclitaxel+carboplatin+bevacizumab with atezolizumab (vs. placebo) followed by maintenance bevacizumab and atezolizumab (vs. placebo) for patients with previously untreated advanced-stage epithelial ovarian-, fallopian tube-, and peritoneal cancer. This global study will enroll approximately 1300 patients both after primary cytoreductive surgery and prior to neoadjuvant therapy per the patient's treatment plan. The co-primary endpoints will be local-assessed progression-free survival (INV-PFS) and overall survival (OS).

3.2 PROGRESSION-FREE SURVIVAL

PFS is defined as the time from randomization to the occurrence of disease progression, as determined by investigators from tumor assessments, per RECIST v1.1, or death from any cause, whichever occurs first. Patients who have not experienced disease progression or death at the time of analysis will be censored at the time of the last tumor assessment. Patients with no post-baseline tumor assessment will be censored on the date of randomization plus 1 day.

INV-PFS will be analyzed using the intent-to-treat population (ITT) population, defined as all randomized patients with patients assigned to treatment groups as per the randomization.

INV-PFS will be tested with the overall Type I error controlled at a two-sided level of 0.05. The null and alternative hypotheses regarding INV-PFS in the ITT population can be phrased in terms of the survival functions $S_A(t)$ and $S_B(t)$ for Arm A (atezolizumab + paclitaxel+carboplatin+bevacizumab) and Arm B (placebo + paclitaxel+carboplatin+bevacizumab), respectively:

$$H_0: S_A(t) = S_B(t) \text{ versus } H_1: S_A(t) \neq S_B(t).$$

INV-PFS will be compared between treatment arms with use of the stratified log-rank test, and the hazard ratio ($HR_{A/B}$) for INV-PFS (with 95% CI) will be estimated using the stratified Cox proportional hazards model. The stratification factors will be those defined by the protocol, and will be based on data collected by the Interactive Voice/Web Response System (IxRS) at the time of randomization [12].

In this study, the co-primary efficacy endpoints are local-assessed progression-free survival (PFS) & overall survival (OS). Therefore, a blinded independent central review is recommended to confirm the local assessment of progression-free survival. The concept of sample-based independent review (i.e., the independent review amongst a sample of

patients) is considered instead of a full review by blinded independent central review (BICR) using the Stone method as described in section 5.

4. AUDIT STRATEGY FOR YO39523

A sample of centrally reviewed cases will be used to decide if a full review of local assessments of progression free survival is needed. The sample will be initiated approximately 24 months after the first patient randomized in the study, 12 months ahead of pre-specified primary analysis of PFS in study protocol, to allow a rapid decision on whether full study review is required. The idea of choosing 24 month from FPI is for implementation consideration, in which the conduct of this sample-based BICR is attached to one of the pre-specified iDMC reviews for safety, eg: every 6 month from FPI. In addition, sufficient time has to be left after this sample-based BICR in any case that a complete full BICR review has to be implemented. Based on the current assumption of enrollment duration of 25 months in total and the yearly dropout rate of 5%, there will be approximately 1248 patients, whose scans will be available for reading. According to the tumor assessment in the protocol, approximately 5720 scans (composite of 40% of total number of scans if patients all finish their scheduled tumor assessments) will be available for BICR. 24 months time point is supported by the operational characteristics (OC) described in section 5.3.

Parexel was contracted by the sponsor to perform the BICR based on a uniform interpretation of radiographic image data for study YO39523. The independent review charter describes the processes, services, and issues related to procedures directed by Parexel for the independent assessment of the radiographic endpoint. In the context of the interim audit strategy, images up to 24 months after the first patient randomized will be collected and analyzed according to Sponsor/Parexel standard procedures.

The essential data of the BICR will be provided by the sponsor to the independent data coordinating center (iDCC) for analysis. The evaluation of the interim audit data will be submitted to the independent data monitoring committee (iDMC) to indicate to the sponsor whether the BICR-assessed progression-free survival interim sample had been accepted according to an approach that adapts a method originally created for defining futility bounds in group sequential designs presented by Stone (2015) [13] where the hazard ratio ratio, the ratio of the hazard ratio of PFS for the treatment effect estimated from the blinded independent central review (BICR) to the corresponding hazard ratio for the local assessments, is used as the metric to define bias (Section 4) without unblinding the treatment assignments or revealing any efficacy results to sponsor.

As to be defined in the independent data monitoring committee charter, when the iDMC issues its recommendation, the Sponsor's data review board makes the final decision as to whether to accept the iDMC's recommendations.

5. AUDIT METHODOLOGY FOR YO39523

5.1 MODEL FREE AUDIT APPROACH

The model free audit approach presented in Stone (2015) is based on an approach to futility analyses developed to be used in group sequential designs, specifically the parameter-free approach as described in Chapter 10 of Jennison and Turnbull. [14]The primary inference of the model-free audit procedure concerns the point estimate for the hazard ratio (HRR) in the full data set, which is equal to the point estimate of BICR-assessed hazard ratio (HR) divided by the point estimate of the local-assessed hazard ratio (HR) . In the model-free audit approach, inference is made about the point-estimates, as opposed to population parameters, as an assessment of bias in the LE in individual studies is made by a comparison of point estimates and not of a statistical comparison of true population HRs between local and BICR for example.

In the model-free approach, absence of bias, or more precisely lack of evidence of meaningful bias, is concluded if there is a low conditional probability that the HRR seen in the random sample of patients would have been observed if, in fact, the point estimate of the HRR in the full trial were unacceptably high.

When applied to this study, the probability of falsely accepting the sample (sensitivity, see section 5.4) will always be 10% if the true HRR in the full study is as small as 1.14. A HRR of 1.14 would occur if the BICR HR was 0.8 and LE HR was 0.7 in the full trial. The size of the sample has then been determined to ensure there is also a high probability (specificity, see section 5.4,) ~90%, of accepting the sample when in fact there is no bias (HRR=1).If the size of the sample changes, the specificity will change accordingly but the sensitivity will stay the same as 90%, which could make sure the regulatory risk is well controlled.

5.2 STATISTICAL PROPERTIES

In the model-free approach, the estimated HRR based on full dataset, HRR_F is defined as:

$$HRR_F = HR_{B,F}/HR_{L,F}$$

Based on its large sample distribution, it could be expressed on the log scale as:

$$\ln(HRR_F) \sim N(\ln(HRR_F), I_F^{-1})$$

where I_F denotes the Fisher information for ln(HRR).

Similarly, after assessment of the sample of the data, we could estimate the HRs based on local evaluations of sampled subjects.

$$\ln(HRR_S) \sim N(\ln(HRR_S), I_S^{-1})$$

Specifically, the two estimates are bivariate normal with means and variance as stated above, and their covariance is I_F^{-1}. It follows that the conditional distribution of $\ln(HRR_S)$ given HRR_F is

$$\ln(HRR_S)|\ HRR_F \sim N(\ln(HRR_S),\ I_S^{-1} - I_F^{-1})$$

Standardized test statistics are defined as

$$Z_F = \ln(HRR_F)\ \sqrt{I_F}\ \text{and}\ Z_S = \ln(HRR_S)\ \sqrt{I_S}$$

Therefore, the conditional distribution of Z_S given Z_F is

$$Z_S|Z_F \sim N(Z_F\sqrt{IS}/\sqrt{IF}\ ,\ (I_F - I_S)/\ I_F)$$

Based on those statistical properties, the threshold for decision making is discussed in the section below.

5.3 THRESHOLD DETERMINATION

When analyzing the sampled data, the following hypothesis testing is specified

H_0 (null hypothesis): $HRR_F \geq HRR_U$

Against

H_1 (alternative hypothesis): $HRR_F < HRR_U$

where

- HRR_F, the estimated hazard ratio ratio based on the full data,
- HRR_U, a maximum acceptable value of the hazard ratio ratio (HRR) on the full data.

The null hypothesis is defined under which the level of bias in local evaluations is unacceptable. If the audit sample leads to rejection of this null hypothesis, we conclude that local evaluations are sufficiently close to BICR (or biased against the experimental treatment) and a full BICR is unnecessary. Please note that these hypotheses concern the final estimate of HRR and compare it with a pre-specified HRR_U. The determination of HRR comes from clinical judgement of bias, eg: how difference the local assessments vary from BICR's review. In this study, we define the HR in BICR review as 0.8, which preserves two-thirds of the pre-specified HR of local assessments in protocol. Therefore, the HRR_U is 0.8/0.7=1.14.

With the model-free approach closed form solutions exist that can be pre-determined prior to data being analyzed so that the properties of the approach are transparent. As specified in [13] and section 5.1, for a 1-sided level α test, we stop and reject H_0 based on the sample of data if:

$$Z_S<\ln(HRR_U)\ \sqrt{I_S}-\varphi^{-1}(1-\alpha)\sqrt{(I_F-I_S)/I_F}]$$

This criterion can be expressed as a bound on the estimate of HRR_S from the data sample

$$HRR_S < \exp[\ln(HRR_U)-\varphi^{-1}(1-\alpha)\sqrt{(I_F-I_S)/I_SI_F}]=HRR_{AT}$$

where

- HRR_S, the estimated hazard ratio based on the sample data,
- HRR_{AT}, the 'acceptance threshold' for the hazard ratio ratio observed in the sample data,
- I_F denotes the Fisher information for logarithm(hazard ratio ratio) in the full data set,
- I_S denotes the information for logarithm (hazard ratio ratio) in the sample data.

  Both I_F and I_S are functions of number of events in both full and sample data from local assessments, the number of events in BICR, and the correlation between BICR and local assessments as given by equations (5) and (6) of Stone et al.
- α is the total type I error for testing H_0, the probability of accepting the sample if bias exist.

Assuming,

- k, the randomization ratio of 1,
- $n_{L,F}$=601, the number of events in full sample size,
- $n_{L,S}$=374, the number of events in audited sampler, based on the current estimation of mPFS and enrollment , and it might come from 1248 enrolled patients
- $r=n_{B,F}/n_{L,F}$=0.782, the ratio of number of events in full sample by the independent committee and local assessments respectively, the ratio is usually less than 1 given the informative censoring
- ρ=0.7, the correlation between BICR and local assessments,
- HRR_u defined as 1.14 to preserve two-thirds of the local-assessed hazard ratio of 0.7 in study YO39523, as stated above

the threshold is calculated as HRR_{AT}=1.06.

Based on the information from three conducted Phase III studies conducted by Genentech/ Roche in ovarian cancers [13,15,16], r is calculated via dividing the number of events reported from BICR by the number of events reported from local assessment. The numeric average of r from those three studies is 0.782.

ρ=0.7, the correlation between BICR and local assessments, is calculated via bootstrap using data from three studies as mentioned above (OCEANS, AURELIA, GOG0218).

Please note that when doing the interim sampling, ρ and r will be adjusted according to calculation from real sample. Therefore, the threshold might be changed as well, so as the specificity. Sensitivity won't change as it does not depend on either the numerical value of threshold nor ρ and r.

5.4 OPERATIONAL CHARACTERISTICS

In the model-free approach, sensitivity and specificity are defined as following:

- Sensitivity - Probability to conclude bias (i.e. observing HRR in the audit sample is greater than the threshold), or more specifically the probability of accepting H_0 when the true HRR=1.14, where H_0 is the presence of bias.
- Specificity - Probability to conclude no bias (i.e. observing HRR in the audit sample smaller than the threshold) , or more specifically the probability of rejecting H_0 when the true HRR = 1.

In all applications of the model-free approach sensitivity is set at 90%, this guarantees that the probability of falsely accepting the sample is 10% in all cases. This then determines the acceptance threshold and specificity for any particular design. The operation characteristics are provided in Table 1, given the determined threshold as 1.06. For example, taking all images up to 24 months after FPI for audit, it's estimated that approximately 1248 patients will be enrolled with approximately 374 events occurred by that time. With 90% sensitivity, , we could make sure that there is only 10% chance to make a wrong call to accept the sample when there is some bias such that the true value of HRR is 1.14. Applying the threshold of 1.06 will give a specificity of 88.3%, i.e. 11.7% chance making a wrong call when there is no bias as the true ratio of BCIR HR to LE HR is 1.

Table 1. The Operational Characteristics of Audit from Interim Sampling

Time-point	Predicted # of enrolled patients	Predicted # of events	Threshold from interim sampling	Sensitivity	Specificity
24 month from FPI	1248	374	1.06	90%	88.3%
	1080	320	1.05	90%	78%
30 month from FPI	1300	469	1.09	90%	99%

5.5 DECISION RULE

If the aforementioned test does not reject H_0 (i.e., $HRR_S \geq HRR_{AT}$), blinded independent central review is conducted for the full set of data. However, if H_0 is rejected (i.e., $HRR_S < HRR_{AT}$), a full blinded independent central review (BICR) will not be conducted is not required.

5.6 IMPLEMENTATION

The sponsor will work with Paraxel in developing a separate document of BICR charter, which defines the scope of logistics activities.

In addition, sponsor will update working order to iDCC vendor (Cytel), who will calculate the statistics HRR at month 24 after first patient in, with the scan reading data sent directly from Paraxel and treatment assignments information from IxRSsystem. After that, iDMC will calculate the statistics and make assessment respect to the threshold accordingly, and send their recommendation to sponsor without revealing the numeric value of HRR. The details could be found in the updated working contract with iDCC vendor and iDMC charter.

6. REFERENCES

1. Zhang JJ et al. Evaluation of blinded independent central review of tumor progression in oncology clinical trials: a meta-analysis. Therapeutic Innovation & Regulatory Science September 7, 2012 and March 2013 vol. 47 no. 2 167-174.

2. Oncologic Drug Advisory Committee—evaluation of radiologic review of PFS in non-hematologic malignancies. cited 2012 Jul 24. Available from: http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/ucm285400.htm

3. Food and Drug Administration (FDA). Summary minutes of the Oncologic Drugs Advisory Committee Meeting, July 24, 2012. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/UCM316323.pdf

4. Dodd et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? J Clin Oncol 2008;26:3791–6.

5. Amit O et al. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. Eur J Cancer 2011;47:1772–8.

6. Dodd et al. An audit strategy for progression-free survival. Biometrics. 2011 September; 67(3): 1092–1099.

7. F. Hoffmann-La Roche Ltd./Genentech, Inc. BLINDED INDEPENDENT CENTRAL REVIEW AUDIT PLAN FOR THE TH3RESA STUDY. 09-April-2013.

8. Zhang et al. Assessment of Audit Methodologies for Bias Evaluation of Tumor Progression in Oncology Clinical Trials. Clin Cancer Res; 2013 May 15;19(10):2637-45.

9. Stone et al. Model free audit methodology for bias evaluation of tumour progression in oncology. Pharmaceutical Statistics, Volume 14, Issue 6 November/December 2015, 455–463.

10. Sun et al. Testing of evaluation bias for progression free survival endpoint in oncology clinical trials. Statist. Med. 2016, 35 3923–3932.

11. F. Hoffmann-La Roche Ltd./Genentech, Inc. *Protocol* YO39523*: A PHASE III, MULTICENTER, RANDOMIZED, STUDY OF ATEZOLIZUMAB VERSUS PLACEBO ADMINISTERED IN COMBINATION WITH PACLITAXEL, CARBOPLATIN, AND BEVACIZUMAB TO PATIENTS WITH NEWLY-DIAGNOSED STAGE III OR STAGE IV OVARIAN, FALLOPIAN TUBE, OR PRIMARY PERITONEAL CANCER.* Version 1.0. 12-Sep-2016.

12. Ruikar 2016 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4763512/

13. Aghajanian et al. OCEANS: A Randomized, Double-Blind, Placebo-Controlled Phase III Trial of Chemotherapy With or Without Bevacizumab in Patients With Platinum-Sensitive Recurrent Epithelial Ovarian, Primary Peritoneal, or Fallopian Tube Cancer. *Journal of Clinical Oncology* 30, no. 17 (June 2012) 2039-2045.

14. Group Sequential Methods: Application to Clinical Trials. Jennison C, Turnbull BW Chapman&Hall

15. Husain et al. Independent radiologic review of AURELIA, a phase 3 trial of bevacizumab plus chemotherapy for platinum-resistant recurrent ovarian cancer. Gynecol Oncol. 2016 Sep;142(3):465-70.

16. Burger et al. Independent radiologic review of the Gynecologic Oncology Group Study 0218, a phase III trial of bevacizumab in the primary treatment of advanced epithelial ovarian, primary peritoneal, or fallopian tube cancer. Gynecol Oncol. 2013 Oct;131(1):21-6.

17. Stone et al. Model free audit methodology for bias evaluation of tumour progression in oncology. Pharmaceutical Statistics, Volume 14, Issue 6 November/December 2015, 455–463.

Appendix 1
Simulation to Compare PFS Audit Strategies

APPENDIX 1:

SIMULATION TO COMPARE PFS AUDIT STRATIGIES

To compare the performance of various PFS audit strategies (including the sampling strategies and the testing methods), we conducted serval simulation studies as presented below.

1.1 SIMULATION SETUP

1.1.1 ERROR MODEL TO SIMULATE PFS DATA

PFS data points in simulated clinical trials were generated as illustrated in Figure A1. This is similar to the simulation procedure used in Amit (2013) and Dodd (2010). We assumed that actual PFS (shown as T in Figure A1) is a continuous variable having an underlying exponential distribution, but they are not directly observed in clinical trials. Instead, disease progression is measured at scheduled assessments. The set of parameters (p1, p2, p3), also referred as error rates, characterizes the probability reporting the PD event earlier than, exactly at, or later than the actual event time (by BICR or LE, in ctrl or trt arm).

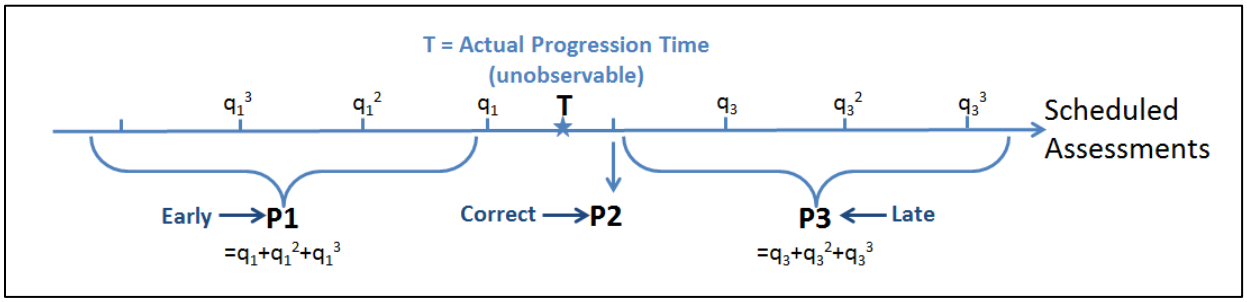


Figure A1: Error Model for Simulation of PFS data

Four sets of parameters (p1, p2, p3), each for (BICR for control arm, BICR for treatment arm, LE for control arm, and LE treatment arm) respectively, introduce the measurement error/variation and potential bias by LE favoring treatment. Magnitude of bias is controlled through setting different error rates between the two arms. In contrast, an unbiased error model will purely characterize measurement error by assigning the same set of error rates to both of the arms. In our simulations, we set error model parameters as below.

BICR (IRF)			
	Early p1	Exact p2	Late p3
Control	0.15	0.7	0.15
Treatment			

LE (INV) w/ no bias			
	Early p1	Exact p2	Late p3
Control	0.2	0.6	0.2
Treatment			

LE (INV) w/ bias			
	Early p1	Exact p2	Late p3
Control	P.bias	0.6	0.2
Treatment	0.2		P.bias

Set p1 = p3 = 0.15, and p2 = 0.7 for BICR, which has no bias and relatively high accuracy 70%. Set p1 = p3 = 0.2, and p2 = 0.6 for LE when there is no bias, but slightly less accuracy. This unbiased LE simulation is used to calculate specificity when comparing various audit strategies.

To introduce evaluation bias, we changed the parameters in the error model for LE to be p1 = p.bias in the control arm and p3 = p.bias in the treatment arm, assuming that LE tend to call progression early in the control arm and late in the treatment arm at the same time. The rest of the error rates remain the same on both arms. Two biased scenarios were simulated: the median biased scenario with p.bias = 0.35, and the high biased scenario with p.bias = 0.65. These two scenarios were used to calculate the sensitivity when comparing various audit strategies.

1.1.2 TRIAL DESIGN PARAMETERS FOR SIMULATION

The true progression time was simulated with a median PFS of 3 months in the control arm, and with a hazard ratio of 0.7 for the treatment arm. Tumor assessment by imaging was equally spaced by every 1.5 month. In addition, we added an optional feature in our simulation to reflect that when a progression is called by LE, there will be no further tumor assessment call that can be made by BICR. This is to mimic the reality that in most clinical trials, tumor image data collection is stopped after a disease progression is called by LE.

1.1.3 SENSITIVITY, SPECIFICITY, AND ROC CURVE FOR STRATEGY COMPARISON

The performance of various audit strategies was evaluated by examining each strategy's Specificity, Sensitivity, and ROC (receiver operating characteristic) curve.

Specificity (Sp) is defined as the probability of declaring LE is reliable when truly there is no bias. It was calculated by the simulation with the error model being set to with no bias for both LE and BICR as described in section 1.1.1. For each audit strategy, there is a test statistics and a threshold. The Sp for the audit strategy was computed as the proportion that the test statistic does not exceed the specific threshold when there is no bias, over one thousand simulated datasets. The higher the Sp is, the better the strategy is.

Sensitivity (Sn) is defined as the probability of detecting evaluation bias (between LE and BICR) when the bias is truly present. It was calculated under two scenarios by simulation with the error model being set to with median or high bias respectively, as described in section 1.1.1. For each audit strategy, the Sn was computed as the proportion that this strategy's test statistic does exceed the specific threshold when there is bias, over one thousand simulated datasets. The higher the Sn is, the better the strategy is.

ROC curve showing Sn as the y-axis vs Sp as the x-axis for each audit strategy was plotted as well. The closer the curve is to the top-right corner, the better the strategy is.

1.2 COMPARING SAMPLING STRATEGY

One challenge for the audit strategies is that when a simple random sample is selected for audit, the audit can only be performed when all image data for the audit patients are available, i.e. after or around databased lock. The Sponsor usually would like to shorten the filing timeline after database lock. In the case of the audit doesn't pass the consistency check between LE and BICR, a full BICR will be needed resulting in a longer timeline and higher cost. To overcome this challenge, an interim sampling strategy is proposed and illustrated in Figure A2. With the interim sampling strategy, if the audit passes the LE-BICR consistence check, further BCIR assessment won't be needed anymore which will save the future work and cost; if the audit doesn't pass the check, BICR can continue as planned, and there is no impact on the original timeline and cost.

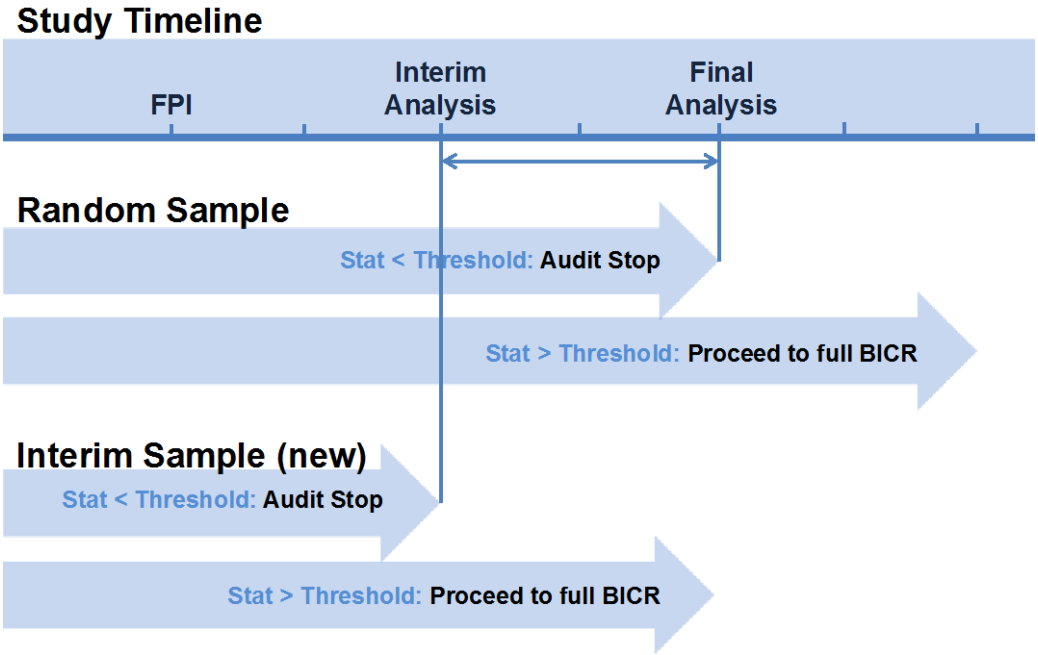


Figure A2. *Comparison of interim and random audit samples in terms of timeline denotes number of subjects in the audit sample.*

To understand how interim sampling performs compared to the usual random sampling,

we set the time of analysis when using the interim audit sample and size of subsample when using the random audit sample in a way such that the total amount of imaging assessments are similar for the two sampling methods. This is to ensure a fair comparison between the two sampling methods by checking their performance based on similar amount of information available.

Figure A3 shows the receiver operating characteristic (ROC) curves using interim and random audit samples, which plots the sensitivity and specificity of the test statistics being compared to a sequence of thresholds as the decision rules. The performance of the testing methods under the median bias scenario was shown is this figure. The left panel applied the PhRMA audit procedure to the interim and random sample, and the right panel used the HRR test statistic. Note that, even though the interim sample has more subjects than the random sample, the number of imaging assessments is similar between the two audit samples. Using either PhRMA or HRR, we see that interim sample performs as well as a random sample with very similar operating characteristics. For scenarios not shown in the figure such as LE evaluation with high bias or with treatment giving larger treatment benefit (i.e. smaller HR), we came to the same conclusion that the operating characteristics are similar for these two sampling procedures for PFS audit. Considering the advantage of no risk of additional cost and timeline delay post database lock but only possible saving of work and cost before database lock, the interim sampling is a generally recommended.

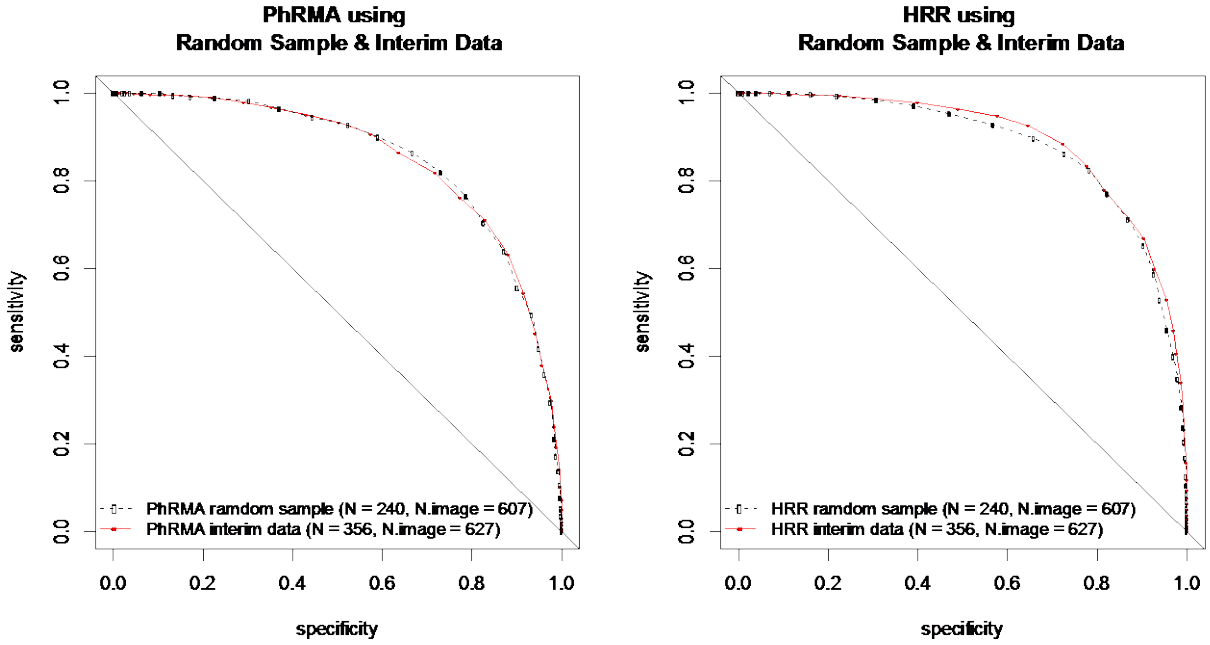


Figure A3. *Comparison of interim and random audit samples in terms of operational characteristics under median bias, using PhRMA and HRR methods. The number of imaging assessments for the two types of audit samples are similar. (N.image is similar).*

1.3 COMPARING TEST STATISTICS

Performance of the three test statistics PhRMA (using the LDR), NCI and HRR, as well as the random and interim sampling methods, was also evaluated in terms of the operational characteristics, i.e., sensitivity and specificity, in both the median and high bias scenarios.

Figure 4A shows the ROC curves of the PhRMA (using LDR), NCI and HRR test statistics in the median bias scenario, using either a random sample presented in the left panel, or an interim sample in the right panel. For the random audit sample, we consider a sample size of either 120 or 240. For the interim sample, 356 patients are audited, with a similar number of imaging assessments as 240 patients in the random sample. The HRR method consistently outperforms the PhRMA and NCI methods for both audit sampling mechanisms and for both audit sample sizes 120 and 240. Additional scenarios not shown in Figure 4A such as with high bias, or with treatment giving larger treatment benefit (i.e. smaller HR), or with other different audit sample sizes, the same pattern is consistently observed. Thus, HRR is generally recommended.

Of note, when using the random sampling mechanism for audit, increasing audit sample size does improve the performance substantially.

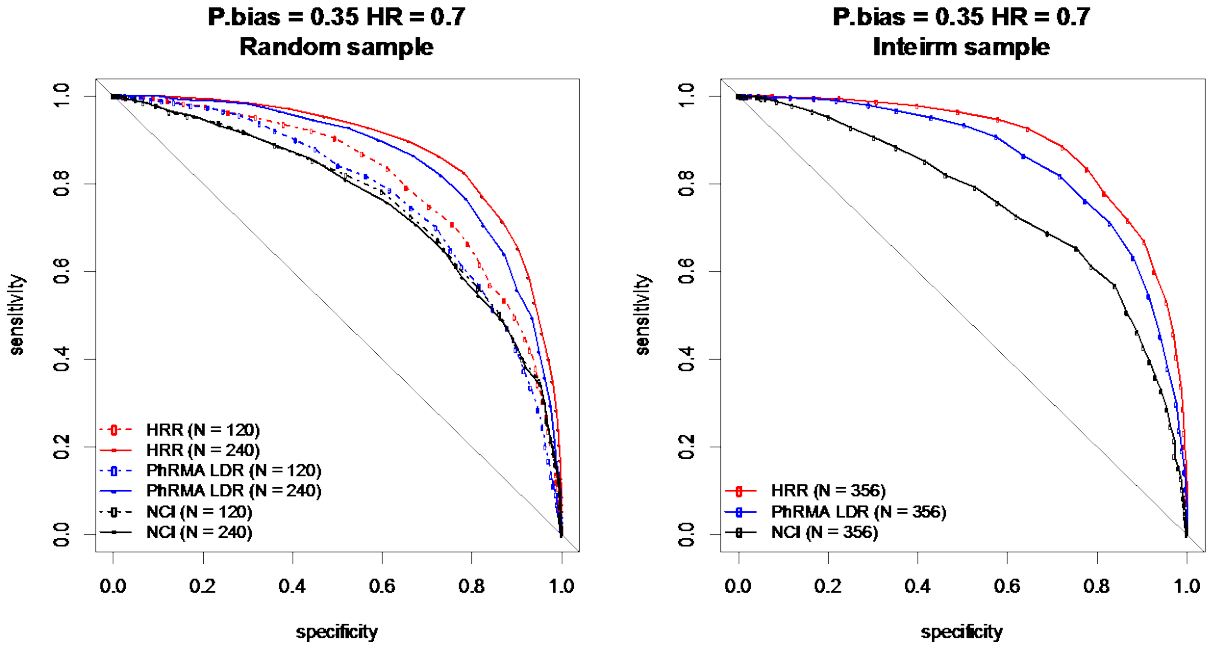


Figure A4. *Comparison of PhRMA, NCI and HRR methods in terms of operational characteristics under median bias, using interim and random audit sample. N denotes number of subjects in the audit sample. Sample sizes of 120 and 240 are considered for random sample, and simple size of 356 is implemented in the interim sample.*

1.4 SUMMARY

Comprehensive simulations based on similar error model used in Amit (2013) and Dodd (2010) were ran to compare the performance of the three test statistics PhRMA (using

the LDR), NCI and HRR, as well as the random vs interim sampling methods. The performance was evaluated based on ROC curve, i.e. sensitivity and specificity along various thresholds for testing. Across various scenarios including when LE with medium or high bias, or with treatment giving larger treatment benefit (i.e. smaller HR), or with different audit sample sizes, consistently we observed the following:

- the interim sampling approach performs as well as the random sampling does
- comparing the PhRMA, NCI, and HRR test statistics, the HRR statistics outperforms the others

Compared to the random sampling approach, the interim sampling has the attractive advantage of no risk of additional cost and timeline delay post database lock but only possible saving of work and cost before database lock.

Therefore, in general, an interim sample-based PFS audit using the HRR method for bias detection is recommended over other options.