STATISTICAL ANALYSIS PLAN

ALKS 4230-007

Study Title:	A Phase 3, Multicenter, Open-Label, Randomized Study of Nemvaleukin Alfa in Combination with Pembrolizumab Versus Investigator's Choice Chemotherapy in Patients with Platinum-Resistant Epithelial Ovarian, Fallopian Tube, or Primary Peritoneal Cancer (ARTISTRY-7)	
Document Date:	Version 3.0	19 Aug 2024
	Version 2.0	05 Feb 2024
	Version 1.0	16 Mar 2023
Based on:	ALKS 4230-007 Protocol v5.0, dated 31 Jan 2024	
Sponsor:	Mural Oncology, Inc. 852 Winter Street Waltham, MA 02451 USA	

CONFIDENTIAL

SIGNATURE PAGE

Author Signature:

[redacted]

[redacted]

Approver Signature:

[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS 6
1. VERSION HISTORY 9
2. INTRODUCTION 10
2.1. Overall Study Design 10
2.2. Study Objectives and Endpoints 12
3. SAMPLE SIZE AND STATISTICAL POWER CONSIDERATION 13
4. GENERAL ANALYSIS DEFINITIONS 14
4.1. Visit Windows 14
4.2. Analysis Population 14
4.2.1. Intent-to-Treat Population 14
4.2.2. Safety Population 14
4.2.3. Pharmacokinetic (PK) Population 15
4.2.4. Pharmacodynamic Population 15
4.3. Definition of Baseline 15
4.4. Subgroup Analyses 15
4.5. General Analysis Methods 16
4.6. Pooling of Centers 17
4.7. Multiple Comparisons/Multiplicity 17
5. SUBJECT INFORMATION 17
5.1. Disposition Information 17
5.2. Demographics and Baseline Characteristics 17
5.3. Protocol Deviations 17
5.4. Medical History 18
5.5. Prior and Concomitant Medications 18
5.6. Prior and Subsequent Anticancer Therapy 18
5.7. Treatment Exposure 18
6. EFFICACY ANALYSES 19
6.1. General Considerations 19
6.2. Analysis of the Primary Endpoint 19
6.2.1. Sensitivity Analyses of the Primary Endpoint 21

6.3. Analysis of Secondary Endpoints 21
6.4. Analysis for Exploratory Efficacy Endpoints 26
7. SAFETY ANALYSIS 26
7.1. General Considerations 26
7.2. Adverse Events 26
7.2.1. Adverse Events of Special Interest 28
7.2.2. Death 28
7.3. Clinical Laboratory Parameters 29
7.4. Vital Signs and ECG 32
7.4.1. Vital Signs 32
7.4.2. Electrocardiograms 32
7.5. Eastern Cooperative Oncology Group (ECOG) Performance Status 32
7.6. Left Ventricular Ejection Fraction 33
8. OTHER EXPLORATORY ANALYSES 33
8.1. Health-Related Quality of Life Analyses 33
8.2. Pharmacokinetic/ Pharmacodynamic Analyses 35
8.2.1. Pharmacokinetic Analyses 35
8.2.2. Blood Pharmacodynamic Analyses 35
8.2.3. Tumor Tissue Pharmacodynamic Analyses 36
8.3. Immunogenicity Analyses 36
9. INTERIM ANALYSES 36
9.1. Interim Futility Analysis of ORR in Each Monotherapy Arm 37
9.2. Interim Efficacy Analysis of OS 37
9.3. Blinding and Independent Data Monitoring Committee 38
10. CHANGES IN CONDUCT OR PLANNED ANALYSES FROM THE PROTOCOL 38
11. DEFINITIONS AND CONVENTIONS FOR HANDLING OF THE DATA 38
11.1. Analysis Visit Windows 38
11.2. Safety Data Handling 39
11.3. Handling of Partial Dates of Concomitant Medication 39
11.4. Reporting Precision 40
12. REFERENCES 40
13. APPENDIX 41

13.1. Extended Time Without an Evaluable Tumor Assessment 41

LIST OF TABLES

Table 1: Abbreviations.....6
Table 2 Version History Summary 9
Table 3: Censoring Rule for Progression-free Survival.....22
Table 4: Clinical Laboratory Assessments 30
Table 5: Eastern Cooperative Oncology Group Performance Status.....32
Table 6: Criteria for Best Overall HRQoL Response 34
Table 7: Pre-Defined Thresholds for Meaningful HRQoL Change.....34
Table 8: Stopping Boundary for Interim Futility Analysis of ORR in Monotherapy Arms 37
Table 9: Summary of Number of Events and Decision Guidance at the Planned OS Analyses.....38
Table 10: Degree of Precision.....40
Table 11: Determination of ≥2 Missed Tumor Assessments.....41

LIST OF FIGURES

Figure 1: Overall Study Design Schematic 11

LIST OF ABBREVIATIONS

Table 1: Abbreviations

Abbreviation	Definition
ACTH	Adrenocorticotropic hormone
ADA	Anti-drug antibody
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
ALK-P	alkaline phosphatase
aPTT	Activated partial thromboplastin time
AST	Aspartate Aminotransferase
AUC	area under the curve
BMI	Body Mass Index
BOR	Best Overall Response
BSA	Body Surface Area
BUN	Blood Urea Nitrogen
CA	Cancer antigen -
CI	Confidence Interval
CLS	Capillary Leak Syndrome
CMH	Cochran-Mantel-Haenszel
CPK	Creatine phosphokinase
CPS	Combined Positive Scores
CR	Complete Response
CRS	Cytokine Release Syndrome
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DCR	Disease Control Rate
DOR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	End of treatment
EWB	emotional well-being

Abbreviation	Definition
eCRF	electronic case report form
ED-5Q-5L	EuroQol 5 Dimension 5-Level
FA	Final analysis
FACT-O	Functional Assessment of Cancer Therapy- Ovarian
FOSI	FACT Ovarian Symptom Index
FWB	functional well-being
GCIG	Gynecologic Cancer InterGroup
HR	hazard ratio
HRQoL	health-related quality of life
IA	Interim analysis
IDMC	independent data monitoring committee
IL	Interleukin
INR	International normalized ratio
irAEs	Immune-related Adverse Events
IRR	infusion-related reaction
ITT	intent-to-treat
IV	Intravenous
LDH	Lactate Dehydrogenase
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic Resonance Imaging
N/A	Not Applicable
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NE	Not Evaluable
NK	Natural Killer [Cells]
ORR	Objective Response Rate
OS	Overall survival
PARPi	poly adenosine diphosphate-ribose polymerase inhibitors
PCS	Potentially clinically significant
PD	Progressive Disease
PD-L1	Programmed Death Ligand-1
PK	Pharmacokinetic(s)

Abbreviation	Definition
PFS	Progression-Free Survival
PLD	Pegylated Liposomal Doxorubicin
PR	Partial Response
PRO	Patient-Reported Outcome
PT	Preferred Term
PWB	physical well-being
RECIST	Response Evaluation Criteria in Solid Tumors
RMST	restricted mean survival time
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
SWB	social/family well-being
SOC	System Organ Class
T3	Triiodothyronine
T4	Free thyroxine
TBILI	Total Bilirubin
TEAE	Treatment-Emergent Adverse Event
TSH	Thyroid-stimulating hormone
TTR	Time to Response
TOI	Trial Outcome Index
ULN	Upper Limit of Normal
VAS	Visual Analogue Scale

1. VERSION HISTORY

This statistical analysis plan (SAP) for Study ALKS 4230-007 is based on the study protocol version 5.0 (dated 31 Jan 2024).

Table 2 Version History Summary

Version	Version Date	Summary of Changes	Rationale
3.0	19 Aug 2024	• Add definition of clinical data cut-off • Clarify definition of baseline to include handling of unscheduled assessments • Reduce number of subgroup analyses • Update duration of exposure algorithm to align with definition of cycle • Clarify algorithm to determine date of last contact for OS • Added primary estimand for OS, PFS, and ORR • Clarify safety summaries to align with ICH requirements • Added analysis of left ventricular ejection fraction • Clarify timing of sponsor unblinding • Correct typos	• Align analysis within the program • Add estimand framework to align with ICH E9(R1)
2.0	05 Feb 2024	• Updated enrollment details to reflect increase in enrollment for Arms 1 and 4 • Update primary endpoint and associated study design and analysis details (sample size, interim	• Update Sponsor to Mural Oncology, Inc • Align with protocol amendment

		analysis, sensitivity analyses) from OS to PFS • Update subgroup analyses • Non substantial updates to clarify analyses	
1.0	16 Mar 2023	Not applicable (N/A)	N/A

2. INTRODUCTION

This statistical analysis plan (SAP) describes the planned statistical methods and data presentation to be used for analyzing and reporting efficacy and safety data for Study ALKS 4230-007. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will be reflected in a protocol amendment.

The primary analysis will include data up to a clinical cut-off date which is determined by the number of events (deaths). The cut-off date is determined once a data extract (before database lock) is available which indicates that the required number of events have occurred. Due to data cleaning activities, the final number of events may deviate from the planned number. The data cut-off date will not be adjusted retrospectively in this case.

2.1. Overall Study Design

This is a Phase 3, multicenter, open-label, randomized study of nemvaleukin in combination with pembrolizumab versus protocol-specific Investigator's choice chemotherapy in patients with platinum-resistant epithelial ovarian, fallopian tube, or primary peritoneal cancer.

Per the original study design, patients were centrally allocated in a randomized fashion (3:1:1:3) to receive either:

- Arm 1: nemvaleukin and pembrolizumab combination therapy
- Arm 2: pembrolizumab monotherapy (closed)
- Arm 3: nemvaleukin monotherapy (closed)
- Arm 4: Investigator's choice chemotherapy. Options for protocol-specific Investigator's choice chemotherapy include one of the following: pegylated liposomal doxorubicin (PLD), paclitaxel, topotecan, or gemcitabine. The Investigator will pre-select the Investigator's choice treatment before the randomization of each patient.

The pembrolizumab monotherapy arm (Arm 2) met its pre-specified futility criteria at the planned futility analysis and was closed to enrollment as of 31 Aug 2023, per the recommendation of the independent data monitoring committee (IDMC). The nemvaleukin monotherapy arm (Arm 3) did not meet its prespecified futility criteria at the planned futility analysis and the IDMC recommended continuation. Under protocol version 4.0, the nemvaleukin monotherapy arm (Arm 3) reached its intended enrollment and thus closed. Because no changes have been made to the objectives related to Arm 3 in protocol version 5.0, the Arm 3 number of

patients was not recalculated. Therefore, in protocol version 5.0, eligible patients will continue to be enrolled in the remaining 2 arms and will be centrally allocated with a randomization ratio of 1:1 to the Combination and Chemotherapy arms (Arm 1 and Arm 4, respectively).

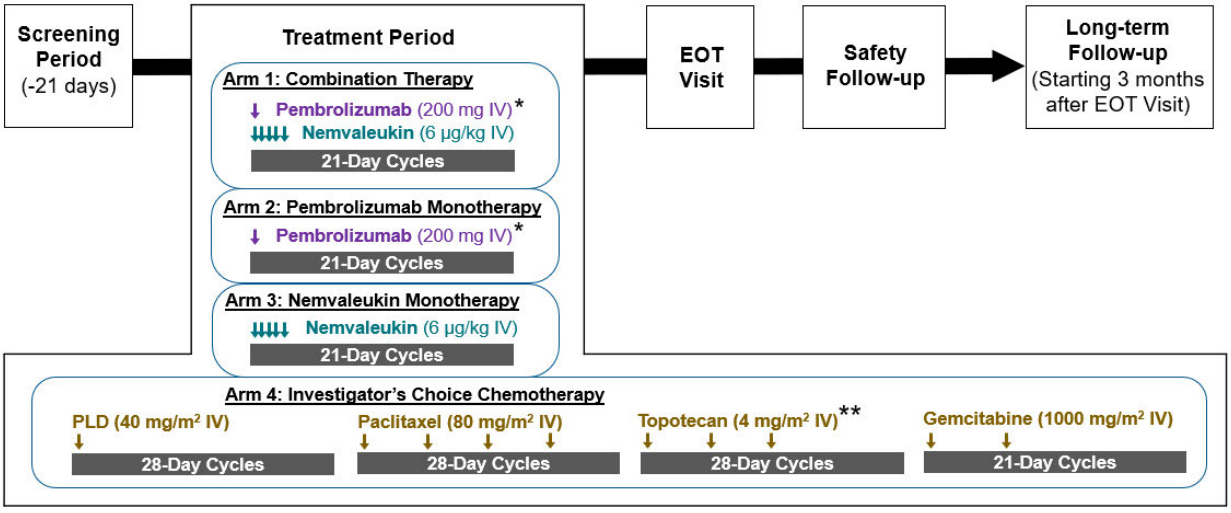
To ensure equal distribution of prognostic factors in the study arms, patients will be stratified according to the following parameters:

- Programmed Death Ligand-1 (PD-L1) status (immunohistochemistry combined positive scores [CPS] ≥10 vs CPS <10)
- Histological subtype (high-grade serous vs non-high-grade serous)
- Investigator's choice chemotherapy (paclitaxel vs other chemotherapies)

Response assessments will include computed tomography (CT) scans and/or magnetic resonance imaging (MRI) every 6 weeks (Year 1) and every 12 weeks (Year 2+), with mandatory relevant imaging of the chest, abdomen, and/or pelvis. Brain imaging should be performed when there is a history of/suspicion of brain metastasis. Response and progression free survival (PFS) will be evaluated per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 guidelines. For patients who complete treatment or discontinue treatment for reasons other than disease progression, every effort should be made to continue monitoring their disease status by radiologic imaging every 6 weeks (Year 1 from start of treatment) or every 12 weeks (Years 2+ from start of treatment) until (1) the assessment of progressive disease, (2) the initiation of new anticancer therapy(ies), (3) withdrawal of consent, (4) death, or (5) end of study, whichever occurs first.

An overall schematic of the study design is provided in Figure 1.

Figure 1: Overall Study Design Schematic



Abbreviations: EOT=end of treatment; IV=intravenous; PLD=pegylated liposomal doxorubicin.
Note: As of protocol version 5.0, the pembrolizumab and nemvaleukin monotherapy arms (Arm 2 and Arm 3, respectively) are closed to enrollment.
*Treatment with pembrolizumab (Arms 1 and 2) is allowed for up to a maximum of 35 cycles (approximately 2 years).
**Alternatively, topotecan may be administered at 1.25 mg/m^2 on Days 1 through 5 of 21-day cycles.

2.2. Study Objectives and Endpoints

The study objectives and corresponding endpoints are summarized below.

Study Objectives	Study Endpoints
Primary Objective: • To evaluate the overall survival (OS) of nemvaleukin in combination with pembrolizumab as compared with chemotherapy in patients with platinum resistant ovarian cancer	**Primary Endpoint:** • OS
Secondary Objectives: • To evaluate the antitumor activity of nemvaleukin in combination with pembrolizumab as compared with chemotherapy • To evaluate the safety of nemvaleukin in combination with pembrolizumab as compared with chemotherapy	**Secondary Endpoints:** • Progression free survival (PFS) as assessed by Investigator, based on Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 • Objective response rate (ORR) as assessed by Investigator, based on RECIST v1.1 • Disease control rate (DCR), duration of response (DOR), and time to response (TTR) as assessed by Investigator, based on RECIST v1.1 • Cancer antigen (CA)-125 response as defined by the Gynecologic Cancer InterGroup (GCIG) • Safety as assessed by treatment-emergent AEs (TEAEs), clinical laboratory parameters, vital signs, and electrocardiograms (ECGs)

Study Objectives	Study Endpoints
Exploratory Objectives: • To evaluate the antitumor activity of pembrolizumab monotherapy; to evaluate the antitumor activity of nemvaleukin monotherapy • To evaluate health-related quality of life (HRQoL) • To determine if any baseline or changes in parameters in tumor tissue and/or peripheral blood may correlate to response to treatment with nemvaleukin in combination with pembrolizumab • To evaluate the pharmacokinetics (PK), immunogenicity, and pharmacodynamic effects of nemvaleukin and/or pembrolizumab in this treatment regimen	**Exploratory Endpoints:** • PFS, ORR, DOR, TTR, and OS of the monotherapy arms as assessed by Investigator, based on RECIST v1.1 • HRQoL as assessed by the following patient-reported outcome (PRO) instruments: the Functional Assessment of Cancer Therapy – Ovarian (FACT-O) questionnaire and the EuroQol 5 Dimension 5-Level (EQ-5D-5L) questionnaire • Pretreatment levels, on-treatment levels, and/or changes from pretreatment levels of the following parameters (in tumor specimens and/or peripheral blood): o Combined positive score (CPS) of programmed death ligand-1 (PD-L1), mutational burden of tumors, and other relevant genetic and non-genetic predictive markers (eg, homologous recombination repair pathway deficiency, breast cancer gene mutation) o Serum concentrations of nemvaleukin (and pembrolizumab as appropriate) o Presence of anti-nemvaleukin antibodies (and anti-pembrolizumab antibodies as appropriate) in serum o Leukocytes including, but not limited to, circulating CD8+ T cells, Tregs, and natural killer (NK) cells o Serum concentrations of interferon-γ, interleukin (IL)-6, and other soluble proteins

3. SAMPLE SIZE AND STATISTICAL POWER CONSIDERATION

The Chemotherapy Arm is considered the comparator arm in this study; the antitumor activity of nemvaleukin in combination with pembrolizumab as compared to that of Investigator's choice chemotherapy in patients with platinum-resistant ovarian cancer will be evaluated.

The primary endpoint was changed from PFS to OS in protocol version 5.0; therefore, the sample size was recalculated based on OS assumptions.

Approximately 366 subjects are planned to be randomized in a 1:1 ratio into the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively). It is expected that approximately 286 deaths will have been observed between the Combination and the Chemotherapy arms at the

final OS analysis. The study has an approximate 85% power to detect an OS hazard ratio of 0.7 (the Combination arm vs the Chemotherapy arm) at an alpha level of 2.5% (one-sided).

The sample size calculation is based on the following assumptions:

1. 2.5% alpha is allocated to the OS endpoint;
2. overall survival follows an exponential distribution with a median of 10 months and 14.3 months in the Chemotherapy arm and Combination arm, respectively;
3. the hazard ratio is 0.7, corresponding to a 4.3 month increase in the median OS;
4. an enrollment period of 25 months; and
5. a yearly dropout rate of 5%.

Under protocol version 5.0, approximately 450 patients are planned, including approximately 366 patients planned across Arm 1 and Arm 4 (approximately 183 patients in each arm). Under protocol version 5.0, no additional patients will be enrolled in the pembrolizumab monotherapy or nemvaleukin monotherapy arms (Arm 2 and Arm 3, respectively).

One planned efficacy interim analysis of OS will be performed when approximately 215 death events (approximately 75% information fraction) are observed in the two arms (the Combination and the Chemotherapy arms). The Lan-DeMets O'Brien-Fleming alpha spending function is constructed to implement group sequential boundaries to control the type I error rate.

The sample size and power calculations were performed in the software R (package "gsDesign").

The monotherapy arms (either pembrolizumab alone or nemvaleukin alone) are included for reference only to examine component effect, and no statistical comparison or testing will be conducted against or between these arms.

4. GENERAL ANALYSIS DEFINITIONS

4.1. Visit Windows

Nominal visit days will be used in the analysis, unless specified otherwise.

4.2. Analysis Population

4.2.1. Intent-to-Treat Population

The Intent-to-Treat (ITT) Population will include all randomized patients regardless of the study drugs (nemvaleukin, pembrolizumab, or Investigator's choice chemotherapy) received.

4.2.2. Safety Population

The Safety Population will include all randomized patients who received any exposure to nemvaleukin, pembrolizumab, or Investigator's choice chemotherapy.

4.2.3. Pharmacokinetic (PK) Population

The PK Population will consist of all patients who received at least 1 dose of nemvaleukin or pembrolizumab and have at least 1 measurable serum concentration of nemvaleukin or pembrolizumab at any scheduled PK timepoint.

4.2.4. Pharmacodynamic Population

The Pharmacodynamic Population will consist of all patients who received at least 1 dose of nemvaleukin, pembrolizumab, or Investigator's choice chemotherapy and have at least 1 available post-baseline pharmacodynamic measurement.

4.3. Definition of Baseline

Baseline is defined as the last non-missing value prior to patient's first dose of study drug.

If an assessment is planned to be performed prior to the first dose of study treatment and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration, unless time of assessment is collected.

Unscheduled assessment will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after the study drug administration. For patients without central laboratory results at baseline, local laboratory results can be used in the determination of baseline.

Study drug is defined as nemvaleukin and/or pembrolizumab, or Investigator's choice chemotherapy.

4.4. Subgroup Analyses

Subgroup analyses will be performed based on the subjects in the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively) for the primary endpoint of OS to assess the consistency of treatment effect across potential or expected prognostic factors. If there are too few OS events in a particular subgroup, ie, <20 events per subgroup level, only descriptive summaries will be provided. Combining relevant subgroup levels may be considered if necessary.

The subgroups include, but are not limited to, the following:

1. Demographic subgroups
    a. Region: North America (NA) and Europe (EU) vs Rest of the World (ROW)
    b. Age: <65 vs ≥65
    c. Race: White vs All Others
2. Biomarker subgroups
    a. Programmed death-ligand 1 (PD-L1) expression at baseline: combined positive score (CPS) <10 vs ≥10
3. Tumor burden subgroups
    a. Visceral vs non-visceral disease

b. Ascites at screening: yes vs no

4. Baseline disease characteristics subgroups

a. Disease Stage at diagnosis: Stage I-III vs Stage IV
b. Baseline Eastern Cooperative Oncology Group (ECOG) Performance Status: 0 vs 1
c. Prior poly adenosine diphosphate-ribose polymerase inhibitors (PARPi): yes vs no
d. Number of prior anticancer therapies: ≤3 vs ≥4
e. Number of prior anticancer therapies in platinum resistant setting: ≤3 vs ≥4
   Platinum resistance is defined as either (1) primary platinum resistance: disease progression 90-180 days following the end of last administered dose of the first line or initial platinum therapy and received at least 4 cycles of platinum ; or (2) secondary platinum resistance: disease progression within 180 days following the end of last administered dose of platinum therapy beyond first-line setting; or (3) platinum refractory: lack of response followed by disease progression while receiving the most recent platinum-based therapy (ie, beyond initial therapy).
f. Investigator's choice of chemotherapy per randomization stratification factor: Paclitaxel vs Other
g. Primary platinum-free interval: <=180 days vs >180 days
   Primary platinum-free interval is defined as the time from the end date of last dose of platinum therapy to the date of progression following the initial or first line of platinum therapy.
h. Platinum response to last platinum therapy: resistant vs refractory
i. Histology type: high-grade serous vs non-high-grade serous (ie, endometrioid carcinoma, clear cell carcinoma, others)

For each subgroup, hazard ratios (HRs) and associated confidence intervals (CIs) will be calculated from an unstratified Cox proportional hazards model for OS. The HRs and two-sided 95% CIs will be presented on a forest plot including the HR and two-sided 95% CI for the overall group. Summaries of the number and percentage of patients experiencing an OS event for each subgroup will be provided along with the median OS by treatment arms for the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively).

4.5. General Analysis Methods

In general, summary statistics (n, mean, standard deviation, median, minimum, and maximum values for continuous variables and number and percentage of patients in each category for categorical variables) will be provided for evaluated variables by treatment arms. Study drug is defined as nemvaleukin and/or pembrolizumab, or Investigator's choice chemotherapy.

Missing data will not be imputed unless otherwise specified.

Multiplicity adjustment will not be applied to sensitivity analyses. The Chemotherapy Arm consists of multiple agents depending on the Investigator's choice; all these agents together will be considered as one Chemotherapy Arm. The statistical comparison or testing will only be conducted between the Combination Arm and the Chemotherapy Arm. No statistical comparison or testing will be conducted against or between the monotherapy arms.

4.6. Pooling of Centers

Data from all study sites will be combined for analysis. Subgroup analysis based on region (NA and EU vs ROW) will be conducted. No analysis for each site is planned.

4.7. Multiple Comparisons/Multiplicity

The overall type I error for this study is strongly controlled at 2.5% (one-sided), fully allocated to OS. The statistical comparison is only performed between the Combination and Chemotherapy arms.

The Lan-DeMets O'Brien-Fleming alpha spending function is constructed to implement group sequential boundaries for OS hypothesis testing to control the type I error.

Efficacy boundaries are based on the assigned type I error rate and the projected number of events at study milestones. The actual boundaries will be determined from the actual number of events at the time of the specified interim analysis using the alpha spending function.

There will be no type I error rate adjustment for the interim futility analysis performed for each of the monotherapy arms since the monotherapy arms are not included for any statistical comparisons.

5. SUBJECT INFORMATION

The subject information analyses will be based on the ITT Population unless otherwise specified.

5.1. Disposition Information

The number and percentage of patients who discontinued the treatment and the reasons for treatment discontinuation will be summarized.

The number and percentage of patients who discontinued the study and the reasons for study discontinuation will be also summarized.

5.2. Demographics and Baseline Characteristics

Demographics and baseline characteristics such as age, race, ethnicity, weight, height, body mass index (BMI), body surface area (BSA), and Eastern Cooperative Oncology Group (ECOG) status will be summarized with descriptive statistics for the ITT population.

BSA will be calculated based on the Du Bois formula as follows: BSA (m^2) = weight $(kg)^{0.425}$ × height $(cm)^{0.725}$ × 0.007184.

5.3. Protocol Deviations

Protocol deviations collected during the study conduct will be categorized and characterized as reportable or non-reportable in accordance with the protocol deviation plan.

Reportable deviations will be summarized and assessed by treatment arms and a supportive listing will be provided.

5.4. Medical History

Medical history information will be coded and summarized for the ITT Population using the most recent version of Medical Dictionary for Regulatory Activities (MedDRA).

Cancer disease history data will be summarized with descriptive statistics for the ITT population.

5.5. Prior and Concomitant Medications

Prior and concomitant medications will be summarized for the ITT population using the World Health Organization-Anatomical Therapeutic Chemical classification system. Prior medications are defined as medications started and ended prior to the first dose of study drug. Concomitant medications are defined as medications, other than study drug, either started prior to first dose and ongoing at the first dose, or started on or after the first dose of study drug. For the summary tables, if a subject has taken a prior or concomitant medication more than once, the subject will be counted only once for the medication.

5.6. Prior and Subsequent Anticancer Therapy

The following data are collected in the electronic case report form (eCRF) for prior anticancer therapy for disease of interest and the new anticancer therapy following study treatment respectively: Name of drug (and/or class), Start and Stop date, Progression date, Best response, and Reasons for discontinuation.

Prior anticancer therapy for disease of interest and the subsequent anticancer therapy will be summarized separately.

5.7. Treatment Exposure

The treatment exposure of study drug (eg, duration of exposure, number of treatment cycles, number of infusions for each cycle, cumulative dose received, relative dose intensity, and missed doses) will be summarized for the Safety Population.

Duration of exposure (weeks) is calculated as (last dose – first dose date + d) ÷ 7, where d is 1 if the participant discontinues study prior to end of the last cycle. For nemvaleukin d is 17. For pembrolizumab d is 21. For PLD, paclitaxel, topotecan, or gemcitabine d is 28, 7, 14, and 14 respectively.

Relative dose intensity (%) is calculated as 100 × [total dose received (μg/kg or mg or mg/m^2) ÷ total protocol planned dose (μg/kg or mg or mg/m^2)]. To calculate BSA based on the Du Bois formula and the actual dose (μg/kg or mg/m^2) in each visit, the weight measured at the first dosing of each cycle will be used unless there is a dose adjustment in which case the weight measured on the dosing date will be used. The dose of nemvaleukin should be adjusted if the patient's weight changes by >10% since the last time the dose was calculated or adjusted. If the weight was missing, the last observed weight from previous visit(s) will be carried forward.

6. EFFICACY ANALYSES

6.1. General Considerations

Antitumor activity and efficacy analyses will be summarized based on the ITT Population, unless specified otherwise.

Analysis of antitumor outcomes (including ORR, DCR, DOR, PFS, and TTR) will be performed based on both Investigator review and central review of the radiographic measurements, respectively, as defined according to RECIST v1.1. The main analysis will be based on investigator assessment and sensitivity analysis will be based on central assessment.

The stratification factors used for randomization (i.e. as entered into the Randomization and Trial Supply Management system) will be applied to all stratified analyses, in particular, the stratified log-rank test (weighted or unweighted), stratified Cox model, and Cochran-Mantel-Haenszel test method. In the event that there are small strata (<5), for the purpose of analysis, strata will be combined to ensure sufficient number of patients, responses, and events in each stratum. The order of stratification factors will be combined in the order of histological subtype, investigator's choice chemotherapy, and PD-L1 status until the count in every stratum is at least 5.

The monotherapy arms (either nemvaleukin alone or pembrolizumab alone) are included in the study for reference only and descriptive estimates of efficacy will be provided for those arms.

6.2. Analysis of the Primary Endpoint

The primary efficacy endpoint is overall survival (OS). OS is defined as the time from randomization to death due to any cause. Patients lost to follow-up before the cut-off date without documentation of death will be censored at the last known alive date.

The date of last contact will be derived for participants not known to have died at the analysis cutoff using the latest complete date among the following:

- Date of randomization
- Participant assessment dates (e.g. blood draws [laboratory, PK], vital signs, ECOG performance status, ECG, tumor assessments, or tumor biopsy dates)
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation
- Adverse Event (AE) start and end dates
- Concomitant medication start and end dates
- Concomitant radiotherapy start and end dates
- Date of contact collected on the long-term follow-up eCRF where status is alive
- Study treatment start and end dates
- Date of last contact and date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up).

Only dates associated with actual contact/assessment of the participant will be used in the derivation. Dates associated with a technical operation unrelated to participant status such as the date a blood sample was processed will not be used. Assessment dates after the cutoff date will not be applied to derive the last contact date.

OS will be calculated as duration from date of randomization to the date of death or censoring.

The statistical hypothesis test, where HR is the hazard ratio, is:

H_0 (Null Hypothesis): HR (combination/chemotherapy) = 1

H_a (Alternative Hypothesis): HR (combination/chemotherapy) < 1

The primary estimand for OS is described by the following attributes:

- Population: patients with platinum-resistant epithelial ovarian, fallopian tube, or primary peritoneal cancer
- Endpoint: OS
- Treatment condition: nemvaleukin in combination with pembrolizumab versus investigator's choice chemotherapy
- Intercurrent event:
    - Start of new anti-cancer therapy is addressed with treatment policy strategy
    - Treatment discontinuation is addressed with treatment policy strategy
- Summary measure: Hazard ratio

The treatment difference in OS will be assessed by the stratified log-rank test, and p-value will be reported. A stratified Cox proportional hazard model will be used for modeling the treatment effect between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively). The Efron method will be used for handling the ties. The HR and its associated 95% CI from the Cox model will be reported.

OS will be displayed by Kaplan-Meier curves. Median OS and its 95% CI will be provided. The CI for median OS will be based on Brookmeyer-Crowley method using log-log transformation (Brookmeyer and Crowley 1982). The 1-year and 2-year OS rate will be estimated using the Kaplan-Meier method. CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation.
Frequency (number and percentage) of patients with an event (death) and censoring reasons will be presented by treatment arm. Censoring reasons are as follows:

- Alive
- Withdrawal of consent
- Lost to follow-up.

Time of Follow-Up for OS

OS follow-up duration is defined as the time from date of randomization to date of death or date which patients were censored for overall survival. Median OS follow-up duration will be estimated using the reverse Kaplan-Meier method, i.e. Kaplan-Meier estimate reversing the OS censoring and event indicators.

6.2.1. Sensitivity Analyses of the Primary Endpoint

The following sensitivity analyses are planned for the primary endpoint:

- OS will be compared between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively) using the unstratified log-rank test based on the ITT Population, and p-value will be reported. The unstratified Cox proportional hazard model will be used to assess the treatment effect between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively) based on the ITT Population. The Efron method will be used for the handling of ties. The hazard ratio (HR) will be reported with 95% CI.
- To account for the potential non-proportional hazards effect associated with immunotherapies, OS will be compared between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively) using the Fleming and Harrington weighted log-rank test with parameter (rho=0, gamma=1) stratified by factors used for randomization based on the ITT Population. The p-value will be reported.
- To address the possible violation of the proportional hazard assumption, a restricted mean survival time (RMST) with the ITT Population based on the area under the curve (AUC) for survival will be implemented using different truncation points (every 3 months starting from the median OS on the Combination arm through a timepoint where at least 5% of patients remain at risk). The estimated RMST difference in OS between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively) will be reported with 95% CIs. The SAS procedure such as RMSTREG will be used with a linear link function, and the model will adjust for the stratification factors.
- Additional analyses of OS adjusting for the effect of subsequent treatment may be performed based on recognized methods [eg, the Rank Preserving Structural Failure Time (RPSFT) model (Robins and Tsiatis 1991)], if a sufficient proportion of participants receive a subsequent therapy. The choice of the method will be based on an examination of the appropriateness of the data to the assumptions required by the method.

6.3. Analysis of Secondary Endpoints

Progression-Free Survival

Progression-free survival (PFS) is defined as the time from randomization to the first documentation of objective tumor progression per RECIST v1.1 or death due to any cause.

For patients who have PD while on study treatment, the date of disease progression will be the date of the first assessment at which PD is objectively documented per RECIST v1.1 . If the patient discontinues study treatment for reasons other than documented PD per RECIST v1.1, then the date of disease progression will be the date of the first assessment at which PD is objectively documented per RECIST v1. during the Follow-up Period and prior to start of new anticancer treatment. Death is considered as a confirmed PD event. In case PD or death are observed immediately after 2 or more missed consecutive tumor assessments, PFS will be

censored at the date of the last tumor assessment immediately before 2 or more missed consecutive tumor assessments. All remaining patients will be censored at the last known date at which the patient was considered progression-free.

PFS will be calculated from the date of randomization to the date of PD or death, whichever occurs first.

The primary estimand for PFS per RECIST v1.1 is described by the following attributes:

- Population: patients with platinum-resistant epithelial ovarian, fallopian tube, or primary peritoneal cancer
- Endpoint: PFS per RECIST v1.1
- Treatment condition: nemvaleukin in combination with pembrolizumab versus investigator's choice chemotherapy
- Intercurrent event:
  - New anti-cancer therapy that started before PD per RECIST v1.1 or death is addressed with hypothetical strategy
  - Treatment discontinuation is addressed with treatment policy strategy
  - 2 or more missed consecutive tumor assessments is addressed with hypothetical strategy
- Summary measure: Hazard ratio

The main analysis for the primary estimand will be based on investigator assessment. Sensitivity analysis will be conducted based on central assessment using similar approach.

A stratified Cox proportional hazard model will be used for modeling the treatment effect between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively). The Efron method will be used for handling the ties. The HR and its associated 95% CI will be reported.

The median PFS will be provided along with the 95% CI. The CI for median PFS will be based on Brookmeyer-Crowley method using log-log transformation (Brookmeyer and Crowley 1982). In addition, Kaplan-Meier curves will be provided. The 6-month, 9-month, or 1-year PFS rate will be estimated using the Kaplan-Meier method. CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation.

The censoring rules for PFS are summarized in Table 3.

Table 3: Censoring Rule for Progression-free Survival

Cases	Analysis
No baseline disease assessments and no death	Censored at the date of randomization
No post-baseline evaluable disease assessments (prior to new anticancer treatment, if initiated) and no death	Censored at the date of randomization

Cases	Analysis
PD per RECIST v1.1 or death documented after ≤1 missed tumor assessment[a]	Progressed at the date of documented PD or death, whichever occurs first
With post-baseline evaluable tumor assessments, new anticancer treatment is not initiated, and no documented PD per RECIST v1.1 or death	Censored at the date of last evaluable tumor assessment
Patients received a new anticancer treatment that started before PD per RECIST v1.1 or death	Censored at the last evaluable tumor assessment where the patient was documented as progression-free on or prior to the intervention
PD per RECIST v1.1 or death at the next tumor assessment immediately after 2 or more missed consecutive tumor assessments[a]	Censored at the date of the last evaluable tumor assessment immediately before 2 or more missed consecutive tumor assessments

Abbreviations: PD=progressive disease; PFS=progression-free survival
[a] Refer to Section 13.1 for details regarding missed tumor assessment.

Frequency (number and percentage) of patients with each event type (PD or death) and censoring reasons will be presented by treatment arm. Censoring reasons are as follows:

- Ongoing in the study without an event
- No baseline assessment
- No adequate post baseline assessments
- Start of new anticancer therapy
- Event after 2 or more missing assessments
- Withdrawal of consent
- Lost to follow-up.

PFS2 (Delayed PFS) Analysis

PFS2 as assessed by the investigator is defined as the time from randomization to subsequent disease progression after initiation of new anticancer therapy, or death from any cause, whichever occurs first. If progression after next-line therapy cannot be measured, a PFS event is defined as end or discontinuation of next-line treatment or death from any cause, whichever occurs first. Patients alive and for whom a PFS event has not been observed should be censored at the last time known to be alive and without second disease progression. The same statistical model for PFS will be used for the PFS2 analysis.

The following censoring rules will be applied for PFS2:

Subjects who did not receive subsequent anti-cancer therapy:

- Subjects who died, the death date is the event date;
- Else the subject's PFS2 is censored at his last known alive date.

Subjects who received subsequent anti-cancer therapy:

- Subjects who had disease progression after the start of subsequent anti-cancer therapy, this disease progression date is the event date;

- Else if a subject died or discontinued subsequent anti-cancer therapy, the \ minimum of (date of death, date of discontinuation of subsequent anti-cancer therapy) is the event date;
- Else the subject's PFS2 is censored at his last known alive date.

Objective Response Rate

Objective response rate (ORR) is defined as the proportion of patients in the analysis population who have a complete response (CR) or partial response (PR) based on RECIST v1.1.

The best overall response (BOR) is the best response recorded from the date of randomization until initial documented disease progression or start of a new anticancer treatment, whichever occurs first, taking into account the requirements for confirmation. Confirmation of CR or PR is required for assessment of BOR unless otherwise specified. The following rules are considered when assigning the confirmed BOR:

- Complete Response (CR) requires at least two determinations of CR at least 4 weeks apart and prior to progression or start of new anticancer therapy. Subsequent documentation of a CR may provide confirmation of a previously identified CR even with an intervening not evaluable (NE), eg, CR NE CR.
- Partial Response (PR) requires at least two determinations of PR or better at least 4 weeks apart (and not qualifying as CR) and prior to progression or start of new anticancer therapy. Subsequent documentation of a PR may provide confirmation of a previously identified PR even with an intervening NE or stable disease (SD), eg, PR NE PR or PR SD PR. However, only 1 intervening NE or SD will be allowed between PRs for confirmation.
- Stable Disease (SD) or non-CR/non-PD (applicable only to participants without measurable disease at baseline) requires at least one SD or non-CR/Non-PD assessment (or better) at least once after the first dose and prior to progression or start of new anticancer therapy at a minimum duration of 6 weeks – 7 days, ie, 35 days (not qualifying as CR or PR).
- Best overall response of progressive disease (PD) requires progression at or prior to the second disease assessment not otherwise qualifying as CR, PR, or SD. Clinical deterioration will not be considered as documented disease progression.
- If the criteria for best overall response of CR, PR, SD, or PD are not met, including due to lack of postbaseline assessments, postbaseline assessments of non-evaluable (NE), start of new anticancer therapy prior to first postbaseline assessment, SD of insufficient duration, or PD after multiple NE assessments, will be consider as NE.

In case there are any post-treatment responses recorded prior to the initiation of new anticancer treatment or progression of disease, whichever occurs first, these responses will be included to determine the BOR. The ORR will be based on BOR.

The primary estimand for ORR per RECIST v1.1 is described by the following attributes:

- Population: patients with platinum-resistant epithelial ovarian, fallopian tube, or primary peritoneal cancer

- Endpoint: ORR per RECIST v1.1
- Treatment condition: nemvaleukin in combination with pembrolizumab versus investigator's choice chemotherapy
- Intercurrent event:
  - Start of new anti-cancer therapy is addressed with while-on-treatment strategy
  - Death prior to response assessment is addressed with composite strategy
  - Treatment discontinuation is addressed with treatment policy strategy
- Summary measure: difference of ORR between nemvaleukin in combination with pembrolizumab and investigator's choice chemotherapy

The main analysis for the primary estimand will be based on investigator assessment. Sensitivity analysis will be conducted based on central assessment using similar approach.

The analysis of ORR between the Combination and the Chemotherapy arms (Arm 1 and Arm 4, respectively) will be conducted by the CochranMantel-Haenszel (CMH) test based on randomization stratification factors. The ORR difference and its associated 95% CI will be reported. A waterfall plot and swimmer plot based on RECIST v1.1 will be used to display the characteristics of the responses in patients.

In addition to ITT population, ORR will be also summarized in the patients who received the study drug and have measurable disease at baseline.

Duration of Response

Duration of Response (DOR) is defined as the time from the first documentation of CR or PR to the first documentation of objective tumor progression or death due to any cause. Patients who never achieve CR or PR prior to starting any new anticancer treatment will be excluded from the analysis. DOR is based on a non-randomized subset of patients (specifically, patients who achieve an objective response). Therefore, a formal hypothesis testing will not be performed for this endpoint.

DOR will be calculated as follows (in weeks): (date of disease progression/death – date of first response (CR or PR) + 1) ÷ 7.

The distribution of DOR will be estimated using the Kaplan-Meier method. Kaplan-Meier estimates for the median DOR will be provided along with two-sided 95% CI. The CI for median will be based on the Brookmeyer-Crowley method using log-log transformation (Brookmeyer and Crowley 1982). Kaplan-Meier curves will also be provided.

The censoring rule for DOR is the same as that of PFS.

Disease Control Rate

Disease control rate (DCR) is defined as the proportion of patients with objective evidence of CR, PR, or stable disease (SD). For SD, measurements must have met the SD criteria at least once after study entry at a minimum interval of 6 weeks – 7 days of visit window, ie, 35 days.

The DCR will be analyzed using the CMH test based on randomization stratification factors. The difference and its associated two-sided 95% CI will be reported.*Time to Response*

Time to response (TTR) is defined as the time from randomization to the first documentation of CR or PR. Patients who never achieve CR or PR prior to starting any new anticancer treatment will be excluded from the analysis. No censoring observation will occur by definition. Time to response is based on a non-randomized subset of patients (specifically, patients who achieve an objective response); therefore, formal hypothesis testing will not be performed for this endpoint. Comparisons between treatment arms will be made for descriptive purposes.

The TTR will be calculated as follows (in weeks): (date of first response [Complete or Partial] – date of randomization + 1) ÷ 7.

Cancer Antigen-125 Response

A CA-125 response is defined as at least a 50% reduction in CA-125 levels from baseline, and the response must be confirmed and maintained for at least 28 days as per the GCIG. The number of responders will be analyzed using the CMH test based on randomization stratification factors. The response difference and its associated two-sided 95% CI will be reported.

In addition, the change from baseline value in CA-125 will be summarized with descriptive statistics.

6.4. Analysis for Exploratory Efficacy Endpoints

The exploratory efficacy endpoints include the PFS, ORR, DOR, TTR, as assessed by Investigator based on RECIST v1.1 and OS for the monotherapy arms. The analysis of these endpoints will be summarized in the same manner as the secondary efficacy endpoints are analyzed, as noted in Section 6.3.

7. SAFETY ANALYSIS

7.1. General Considerations

All safety endpoints will be summarized for the Safety Population.

In general, all by-visit summaries of safety parameters will only include baseline, Day 1 of first 6 cycles, and the EOT visit. Additional summaries will be derived as the minimum and maximum for all on-study assessments and presented similarly at the end of the summary table. Unscheduled measurements will not be used in computing the descriptive statistics for change from baseline at each post-baseline time point. However, the maximum and minimum calculations will use all post-baseline data, including any unscheduled assessments.

7.2. Adverse Events

Adverse events will be coded using the most recent version of MedDRA. Treatment-emergent AEs (TEAEs) are defined as AEs that that have an onset on or after the day of the first dose of study drug until 30 days after the last dose of study drug (90 days for serious adverse events [SAEs]). or until the start of new anti-cancer therapy, whichever is earlier. A single post-

treatment time window will be used for SAE in the TEAE definition for consistency across treatment arms.

Events where the onset date was the same as the study drug start date will be assumed to be treatment-emergent, unless the study drug start time and the AE start time are collected, and the AE start time is prior to the study drug start time. If an incomplete onset date is collected for an AE, the AE will be assumed to be treatment-emergent unless there is evidence that confirms that the AE was not treatment-emergent (eg, the AE end date is prior to the date of the first dose of study drug).

TEAEs will be summarized by preferred term (PT) and by system organ class (SOC) when applicable. Patients reporting more than one AE within a SOC will be counted only once for that SOC. Patients reporting the same AE more than once will be counted only once for that PT.

The severity of TEAEs will be assessed by National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE). For TEAEs that are not specified in the NCI CTCAE, severity grade is assigned by the criteria in the protocol. If a patient has the same AE on multiple occasions, the highest severity recorded for the event will be presented in the AEs-by-severity table.

An overview table, including the number and percentage of patients with TEAEs, serious TEAEs, serious treatment-related TEAEs, TEAEs related to study drug, grade 3 to 4 TEAEs, grade 3 to 4 treatment-related TEAEs, TEAEs leading to treatment discontinuation, TEAEs leading to dose reduction, TEAEs leading to dose interruption, and TEAEs leading to death, will be summarized.

In addition, the number and percentage of subjects experiencing TEAEs will be summarized by treatment arm for the following:

- TEAEs by SOC and PT
- TEAEs by PT
- Treatment-related TEAEs by SOC and PT (Related to Any Study Treatment)
- Treatment-related TEAEs by PT (Related to Any Study Treatment, and related to each drug component for combination therapy)
- TEAEs by maximum CTCAE toxicity grade and PT
- Treatment-related TEAEs by maximum CTCAE toxicity grade and PT (Related to Any Study Treatment, and related to each drug component for combination therapy)
- Serious TEAEs by SOC and PT
- Serious TEAEs by PT
- Treatment-related Serious TEAEs by PT (Related to Any Study Treatment, and related to each drug component for combination therapy)

- TEAEs leading to discontinuation by PT (of Any Study Treatment, and of each drug component for combination therapy)
- TEAEs leading to dose interruption by PT (of each drug component as applicable)
- TEAEs leading to dose reduction by PT (of each drug component as applicable)
- TEAEs leading to death by PT

Listings will be provided for all AEs, serious TEAEs, AESIs, TEAEs leading to treatment discontinuation, TEAEs leading to dose reduction, TEAEs leading to dose interruption, and TEAEs leading to death. Adverse events observed beyond 30 days (90 days for SAE) after the last dose of study drug will be listed. The AE onset date and duration will be shown on all listings of AEs.

The following Preferred Terms will be grouped in order to describe a medical condition rather than a single event: neutropenia (neutropenia/ neutrophil count decrease), anemia (anemia/ haemoglobin decrease), thrombocytopenia (thrombocytopenia/ platelet count decrease), lymphopenia (lymphopenia/ lymphocyte count decrease), and supraventricular extrasystoles (supraventricular extrasystoles/ supraventricular tachycardia).

Summaries of characteristics for neutropenia will be provided.

7.2.1. Adverse Events of Special Interest

Adverse events of special interest (AESIs) are known or potential risks following nemvaleukin administration for which additional guidance is provided. These include infusion-related reaction (IRR)/Cytokine Release Syndrome (CRS), Capillary Leak Syndrome (CLS), and Immune-related Adverse Events (irAEs).

The number and percentage of patients experiencing at least one TEAE for AESIs as reported by the investigator per the AESI eCRF will be summarized for each AESI category as follows:

- Overview of AESI, along with characteristics for each AESI category
- AESIs by AESI category and PT
- AESIs by AESI category, maximum CTCAE toxicity grade and PT

7.2.2. Death

The death reasons will be collected in the electronic case report form (eCRF) and will include: 1) Progressive Disease, 2) Adverse Event, 3) Other, or 4) Unknown.

The number and percentage of patients for each death reason will be summarized for all participants and participants who died during treatment period respectively.

Listings will be provided for patients who died during the study for any reason, and for patients who died due to an AE.

7.3. Clinical Laboratory Parameters

Blood and urine samples for clinical laboratory assessments will be collected at pre-specified timepoints. Specific hematology, biochemistry, urinalysis, and coagulation assessments are listed in Table 4.

Table 4: Clinical Laboratory Assessments

Hematology	Biochemistry		Urinalysis	Coagulation
• Hematocrit • Hemoglobin • Red blood cell count • Total and differential (absolute) white blood cell count • Absolute counts for neutrophils, lymphocytes, monocytes, and eosinophils • Platelet count • C-reactive protein	<u>General Chemistry</u> • Albumin • Bicarbonate • Calcium • Chloride • Creatine phosphokinase (CPK) • Glucose • Magnesium • Phosphorus • Potassium • Sodium • Total protein • Uric acid • Amylase • Lipase <u>Renal Function Tests</u> • Blood urea nitrogen (BUN) • Creatinine	<u>Liver Function Tests</u> • Alanine transferase (ALT) • Alkaline phosphatase (ALK-P) • Aspartate aminotransferase (AST) • Lactic dehydrogenase (LDH) • Total bilirubin <u>Endocrine Tests</u>[a] • Thyroid-stimulating hormone (TSH)[b] • Adrenocorticotropic hormone (ACTH) • Total cortisol • Free thyroxine (T4) • Triiodothyronine (T3)	• Color and appearance • pH • Specific gravity • Ketones • Protein • Glucose • Bilirubin • Nitrite • Urobilinogen • Leukocytes • Occult blood • Microscopic examination of sediment, *only if urinalysis dipstick results for blood, leukocytes or protein are abnormal* (ie, 2+ or higher)	• International normalized ratio (INR) • Activated partial thromboplastin time (aPTT)

[a] Endocrine tests will be performed for all patients at Screening. On-treatment endocrine tests are only applicable to Arm 1, 2, and 3.

[b] If TSH result is abnormal, collect total T3 or free T3 and Free T4.

Laboratory parameters will be presented in international standard units. Scheduled laboratory parameters assessed by central laboratory will be included in the laboratory results summaries by visit timepoints, unless specified otherwise. Unscheduled assessments as well as assessments by local laboratories will be included in worst case post baseline assessments.

Results of clinical laboratory tests will be summarized at pre-specified visits (baseline, Day 1 of first 6 cycles, EOT visit, as well as worst case post-baseline) by treatment arm for the actual values and for change from baseline.

In addition, hematology and chemistry laboratory determinations will be categorized according to NCI CTCAE grades and shifts from baseline NCI CTCAE grades to maximum and final post-baseline grades will be summarized for tests that are gradable by CTCAE. The baseline and final grades will be defined respectively as the grade of the last measurement collected prior to the first dose of study drug, and as the last post-baseline measurement collected no more than 30 days after the last dose of study drug. The maximum NCI CTCAE toxicity grade value will be the value with the highest NCI CTCAE toxicity grade collected after the first dose of study drug and within 30 days after the last dose of study drug. In cases where multiple values are collected on the same day, the maximum grade value will be selected as the value for that day.

For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia separately. If a lab parameter is graded in both directions (e.g. glucose: hyperglycemia and hypoglycemia), then low direction toxicity grades at baseline and post baseline will be set to 0 when the variables are derived for summarizing high direction toxicity, and vice versa.

Additional laboratory results that are not part of NCI CTCAE will be presented according to the categories: Low (below normal limit), Normal (within normal limits) and High (above normal limit) according to the laboratory normal ranges. Furthermore, only the numeric part in laboratory values that contain non-numeric qualifiers, such as less than (<) a certain value, or greater than (>) a certain value, will be used in the summary statistics.

Detailed listings of data for patients experiencing NCI CTCAE grade ≥3 hematology and chemistry values will be provided. All measurements collected after the first dose of study drug and within 30 days after the last dose of study drug will be included in these listings.

Number of patients who meet the following criterion will be summarized:

- ALT > 3× upper limit of normal (ULN), ALT > 5× ULN, ALT > 10× ULN, ALT > 20× ULN
- AST > 3× ULN, AST > 5× ULN, AST > 10× ULN, AST > 20× ULN
- (ALT or AST) > 3× ULN, (ALT or AST) > 5× ULN, (ALT or AST) > 10× ULN, (ALT or AST) > 20× ULN
- Total Bilirubin (TBILI) > 1.5× ULN, TBILI > 2× ULN
- ALP > 1.5× ULN
- Concurrent ALT > 3× ULN and TBILI > 1.5× ULN
- Concurrent AST > 3× ULN and TBILI > 1.5× ULN
- Concurrent (ALT or AST) > 3× ULN and TBILI > 1.5× ULN
- Concurrent ALT > 3× ULN and TBILI > 2× ULN
- Concurrent AST > 3× ULN and TBILI > 2× ULN

- Concurrent (ALT or AST) > 3× ULN and TBILI > 2× ULN
- Concurrent (ALT or AST) > 3× ULN and TBILI > 2× ULN and ALP > 2× ULN
- Potential Hy's law: Concurrent (ALT or AST) > 3× ULN and TBILI ≥ 2× ULN and ALP ≤ 2× ULN

Concurrent measurements are those occurring within 3-day window.

A scatter plot of maximum total bilirubin versus maximum ALT/AST will be provided. A listing will be provided for patients who meet potential Hy's law.
A listing will be provided for patients who experienced pregnancy during the study.

7.4. Vital Signs and ECG

Vital signs (eg, blood pressure, pulse, respiratory rate, and body temperature), height, and weight as applicable will be assessed at pre-specified timepoint(s).

A 12-lead ECG will be conducted at pre-specified timepoint(s).

7.4.1. Vital Signs

Descriptive statistics for vital signs and changes from baseline values after the first dose of study drug will be presented at pre-specified timepoints (baseline, Day 1 of first 6 cycles, EOT visit, as well as post-baseline minimum and maximum) by treatment arm.

Patients who meet the potentially clinically significant (PCS) criteria in vital signs will be summarized by treatment arm. The PCS is defined as SBP ≥160 or DBP ≥100, or SBP ≤90 or DBP ≤60, or Temperature >40°C.

7.4.2. Electrocardiograms

Frequency (number and percentage) of patients with post-baseline ECGs that are deemed clinically significant by the Investigator will be summarized. In addition, a detailed data listing for patients with ECG of PCS will be provided.

7.5. Eastern Cooperative Oncology Group (ECOG) Performance Status

The ECOG Performance Status assesses the patients' activity status. Descriptions of activity status are presented in Table 5. Possible scores are 0 to 5. The number and percentage of patients with baseline score versus post-baseline scores will be summarized by treatment arms in a shift table.

Table 5: Eastern Cooperative Oncology Group Performance Status

Grade	Description
0	Normal activity. Fully active, able to carry on all pre-disease performance without restriction.
1	Symptoms, but ambulatory. Restricted in physically strenuous activity, but ambulatory and able to carry out work of a light or sedentary nature (eg, light housework, office work).

Grade	Description
2	In bed <50% of the time. Ambulatory and capable of all self-care, but unable to carry out any work activities. Up and about more than 50% of waking hours.
3	In bed >50% of the time. Capable of only limited self-care, confined to bed or chair more than 50% of waking hours.
4	100% bedridden. Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair.
5	Dead

7.6. Left Ventricular Ejection Fraction

Left ventricular ejection fraction will be performed only when pegylated liposomal doxorubicin was selected as physician's choice of chemotherapy. For subjects received pegylated liposomal doxorubicin, descriptive statistics for left ventricular ejection fraction and changes from baseline values will be presented. Frequency (number and percentage) of participants meeting clinically significant criteria (shift from normal at baseline to >=20%, >=15% decrease from baseline) will be summarized.

8. OTHER EXPLORATORY ANALYSES

8.1. Health-Related Quality of Life Analyses

The functional assessment of cancer therapy-ovarian (FACT-O) is a 39-item validated questionnaire about the past 7 days using a 5-point Likert-type scale (Basen-Engquist et al, 2001). It includes 5 subscale domains.

The first 4 domains include a set of general items that assess physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), and functional well-being (FWB); these items are not specific to one type of cancer. The fifth module is comprised of 12 items that are specific to women with ovarian cancer. In addition to scores from these 5 domains, three other summary scores will be calculated: 1) the FACT-O Trial Outcome Index (TOI) calculated as the sum of the ovarian subscale scores, the PWB, and FWB scores; 2) the FACT/National Comprehensive Cancer Network (NCCN) Ovarian Symptom Index (FOSI), a validated score based on 8 items selected from the FACT-O; 3) the FACT-O total score, calculated as the sum of all 39 items. All FACT-O scores will be calculated according to the guidelines from the instrument developers, which can be found at https://www.facit.org/measures/FACT-O. The FACT-O will be administered to all study patients on Day 1 of every cycle through Year 1, every 12 weeks thereafter, EOT visit and the 30-Day Safety Follow-up visit. In addition, patients randomized to Arms 1 and 3 will complete the FACT-O on Day 5 of cycles 1 and 2.

The EQ-5D-5L is a validated quality of life questionnaire developed by the EuroQol Group in order to provide a simple, generic utility measure for characterizing current health states of patients. The EQ-5D-5L is designed for self-completion by patients. It consists of 2 parts – the EQ-5D-5L descriptive system and the Visual Analogue Scale (VAS).

The EQ-5D-5L descriptive system comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels.

The 5-dimensional, 5-level systems are converted into an index value. Values for theoretically possible health states are calculated using a regression model and weighted according to the social preferences of the US population (Van Hout et al, 2012).

The VAS records the respondent's self-rated health on a vertical VAS. The VAS 'thermometer' has 100 (best imaginable health state) at the top and 0 (worst imaginable health state) at the bottom. This information can be used as a quantitative health state measure as judged by the individual respondents. During the long-term follow up period, only the VAS will be administered to patients in all treatment arms.

The HRQoL endpoints will be analyzed based on the ITT Population who have a baseline and at least one post-baseline assessment (for FACT-O and EQ-5D-5L separately). Compliance rate will be determined by the number of evaluable forms divided by the number of expected forms. Evaluable forms are those where ≥50% of questions for each domain were answered.

Based on pre-defined within-person meaningful change, the proportion of patients with a best response of 'improved', 'no change' and 'worsened' will be calculated for FACT-O TOI, FOSI and FACT-O Total. Best response will be derived based on evaluable HRQoL data collected from randomization up to the earliest of progression or death. Criteria for determining best HRQoL response are shown below.

Table 6: Criteria for Best Overall HRQoL Response

Category	Criteria
Improved	Two visit responses of "improved" a minimum of 21 days apart without an intervening visit response of "worsened".
No Change	Does not qualify for overall score response of "improved". Two visit responses of either "no change" or "improved" and "no change" a minimum of 21 days apart without an intervening visit response of "worsened".
Worsened	Does not qualify for overall score response of "improved". A visit response of "worsened" without a subsequent response of "improved" or "no change" within 21 days.
Other	Does not qualify for one of the above.

Table 7: Pre-Defined Thresholds for Meaningful HRQoL Change

	Category		
Scale	**Improved**	**No Change**	**Worsened**
TOI	≤ −7	> −7 and <7	≥7
FOSI	≤ −3	> −3 and <3	≥3
FACT-O Total	≤ −9	> −9 and <9	≥9

The absolute values and changes from baseline at each administration timepoint in total score and subscale domain scores for the FACT-O and EQ-5D-5L index and VAS scores will be

descriptively summarized. The number and percentage of patients in the best response will be descriptively summarized.

The change from baseline in scores from FACT-O scores (TOI, FOSI, and FACT-O total), and EQ-5D-5L index value and VAS will be analyzed using a mixed model with repeated measurements with an unstructured variance-covariance matrix based on the observed data. The model will include stratification factors, treatment arm, and interaction term of visit and treatment as categorical variables, and baseline score as a covariate. The Kenward-Roger approximation will be used to adjust the denominator degree of freedom. The change from baseline in the average of all post-baseline visits in the combination arm will be compared to that of the chemotherapy group.

Additional analysis for HRQoL endpoints will be specified in a separate Health Economics and Outcomes Research SAP.

8.2. Pharmacokinetic/ Pharmacodynamic Analyses

The PK and pharmacodynamic analyses will be conducted separately by treatment arms on the PK/pharmacodynamic population as appropriate.

8.2.1. Pharmacokinetic Analyses

Serum samples for evaluation of nemvaleukin PK and/or pembrolizumab PK will be obtained from each patient at pre-specified timepoints.

Concentration data will be summarized descriptively by treatment arm as appropriate according to nominal (protocol specified) sampling times. PK concentration data from this study may be used in a subsequent population PK analysis or other *post hoc* analyses conducted outside of this study.

8.2.2. Blood Pharmacodynamic Analyses

The pharmacodynamic effect of nemvaleukin will be assessed by measuring circulating CD8+ T cells, T_{regs}, and NK cells in peripheral blood from each patient at pre-specified timepoints. Expression of interleukin (IL)-2 receptors on immune cells and T cell receptor clonality may also be assessed.

In addition, serum samples will be obtained from each patient at pre-specified timepoints. Concentration of multiple pro-inflammatory cytokines including interferon-γ, tumor necrosis factor-α, IL-1β, IL-6, IL-10, and other soluble proteins will be determined.

Circulating tumor DNA may be isolated from plasma collected at pre-specified timepoints and may be subject to genetic and epigenetic analyses.

Pharmacodynamic data will be summarized descriptively, including absolute cell counts or proportion and concentrations at nominal (protocol-specified) timepoints, as well as fold change from baseline for circulating leukocytes (including T and NK cells) and cytokines (including interferon-γ and IL-6). Additional analyses may be conducted as deemed necessary upon exploration of data.

8.2.3. Tumor Tissue Pharmacodynamic Analyses

Tumor biopsy samples from patients will be processed and may be analyzed by immunohistochemistry and/or immunofluorescence for tumor infiltrating immune cells. The baseline (pre-treatment) values, post-treatment values, and the changes from baseline in density of immune cells in tumor tissues may be summarized descriptively in tables.

Relevant markers for disease (eg, breast cancer gene status and homologous recombination repair pathway deficiency status, tumor mutational burden scores, and PD-L1 expression) may be summarized descriptively. Additional analyses may be conducted as deemed necessary upon exploration of data.

8.3. Immunogenicity Analyses

Immunogenicity endpoints include the following:

1. Presence of anti-drug antibodies (ADAs)
2. Titer of ADAs
3. Frequency of the presence of anti-drug antibodies by relevant treatment arms

Time of onset of immunogenicity and time of resolution of immunogenicity will be calculated and summarized. Time of onset is defined as the first timepoint with a positive result for ADA; time of resolution is defined as the first timepoint when the previously observed positive for ADA returns to negative for ADA. Duration of positive immunogenicity response (ADA duration) is calculated as the difference between time of resolution of immunogenicity and time of onset of immunogenicity.

Summaries of baseline ADA status (positive or negative) along with number and percentage of participants who have treatment emergent ADA will be provided. Treatment emergent is defined as either (1) negative at baseline and positive post baseline or (2) positive at baseline and positive post baseline where the titer levels demonstrate a 4-fold increase post baseline.

Positive samples titer values will be categorized as appropriate (eg. <100, ≥ 100 to <1000, >1000) and summarized.

Additional immunogenicity characterization (eg, presence of neutralizing anti-drug antibodies) will be summarized.

9. INTERIM ANALYSES

The Sponsor's study team members will be blinded to any aggregate summaries of efficacy data by assigned treatment group during the study conduct. The interim analyses planned in this study are described below. All of the analyses will be conducted by an independent statistician under the guidance of an independent data monitoring committee (IDMC). More details can be found in the IDMC charter. Upon review of the IA by the IDMC and sponsor senior management, the study team may be unblinded after the IA to facilitate regulatory submission.

9.1. Interim Futility Analysis of ORR in Each Monotherapy Arm

The monotherapy arms (either nemvaleukin alone or pembrolizumab alone) are included in the study as reference arms to isolate the component effect; thus, no statistical comparison or testing will be conducted against or between the monotherapy arms.

Interim analyses will be conducted separately when the 24th patient enrolls into the nemvaleukin monotherapy arm and the 12th patient enrolls into the pembrolizumab monotherapy arm; these analyses will be conducted by an independent statistician. The interim analysis on the ORR will provide a statistical recommendation for early stopping for futility.

The stopping boundary is summarized in Table 8.

Table 8: Stopping Boundary for Interim Futility Analysis of ORR in Monotherapy Arms

Monotherapy Arm:	Interim analysis at:	Recommend stopping enrollment if:
Pembrolizumab Monotherapy Arm	First 12 dosed patients	0 or 1 response (confirmed PR or CR)
Nemvaleukin Monotherapy Arm	First 24 dosed patients	0 response (confirmed or unconfirmed PR or CR or durable SD lasting for ≥3 months)

Abbreviations: CR=complete response; PR=partial response; SD=stable disease.

If the futility criteria are met, further enrollment into the arm will stop; patients in the arm receiving benefit (SD or better) may remain on study until progression. Crossover to another treatment arm is not allowed. (Note: all patients at any time who progress will be immediately discontinued from treatment.)

There will be no type I error rate adjustment caused by this futility analysis, as the monotherapy arms are not included for any statistical comparisons.

9.2. Interim Efficacy Analysis of OS

One interim analysis (IA) of OS are planned in the study. The statistical comparison will be performed between the Combination and the Chemotherapy arms.

The study is event-driven. The planned OS IA will be performed when approximately 215 events (at approximately 75% information fraction) are observed in the two arms (the Combination and the Chemotherapy arms). The final analysis of OS will be performed when approximately 286 deaths will have been observed in the two arms (the Combination and the Chemotherapy arms). The Lan-DeMets O'Brien-Fleming alpha spending function is constructed to implement group sequential boundaries to control the overall type I error rate at 2.5% (one-sided) for OS.

Table 9 summarizes the analysis strategies. If the actual number of OS events at the IA and FA differs from those specified in the table, the boundaries will be adjusted using the Lan-DeMets O'Brien-Fleming spending function accordingly.

Table 9: Summary of Number of Events and Decision Guidance at the Planned OS Analyses

Analysis	Expected Number of Events at the Time of Analysis	Efficacy Boundary	
		p-value[a]	HR
OS IA	~215 OS events in the Combination and the Chemotherapy arms	0.0096	0.7265
OS FA	~286 OS events in the Combination and the Chemotherapy arms	0.0221	0.7882

Abbreviations: FA=final analysis; HR=hazard ratio; IA=interim analysis
[a] p-value (1-sided) is the nominal significance level for testing and will be used to claim crossing of a boundary.

9.3. Blinding and Independent Data Monitoring Committee

An IDMC will be established by the Sponsor to analyze and interpret the data for the interim analysis. Members will include experts in oncology and biostatistics who are not participating in this study in any capacity and do not have any external conflict of interest for the study. The IDMC members and specific duties will be fully described in an IDMC charter.

The Sponsor's study team members will remain blinded to any aggregate summaries of efficacy data by assigned treatment group during the study conduct until the primary efficacy analysis timepoint. Upon review of the IA by the IDMC and sponsor senior management, the study team may be unblinded after the IA to facilitate regulatory submission.

10. CHANGES IN CONDUCT OR PLANNED ANALYSES FROM THE PROTOCOL

There are no changes in conduct or planned analyses from the protocol.

11. DEFINITIONS AND CONVENTIONS FOR HANDLING OF THE DATA

11.1. Analysis Visit Windows

Scheduled analysis visits are visits at scheduled timepoints as specified in the protocol.

Scheduled analysis visits during the study period will be the same as the nominal visits collected in the eCRF. There will be one valid value of assessment kept for each scheduled analysis visit in summary/analysis statistics.

Unscheduled visits are visits with data collected outside of scheduled time points. Unscheduled visits will not be used for by-visit summary/analysis statistics, unless specified otherwise.

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

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

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

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

11.2. Safety Data Handling

All efforts should be made to obtain missing information from the Investigator. For vital signs, laboratory testing (chemistry, hematology, urinalysis), and 12-lead ECGs, only observed data will be used for analyses, and missing data will not be imputed.

Adverse events with missing severity are classified as "severe"; AEs with missing relationship to study drug are classified as "related".

Missing start dates of AEs will be imputed following the algorithms below:

- If the day, month, and year are all missing:
  - Use the date of the first study drug administration.
- If the day and month are missing:
  - Use the date of the first study drug administration if the year is the same as the year of the first study drug administration.
  - Use the day and month of 31 December if the year is prior to the year of the first study drug administration.
  - Use the day and month of 01 January if the year is after the year of the first study drug administration.
- If the day is missing:
  - Use the date of the first study drug administration if the month and year are the same as those of the first study drug administration.
  - Use the last day of the month if the month is prior to and the year is the same as those of the first study drug administration.
  - Use the first day of the month if the month is after and the year is the same as those of the first study drug administration.
  - For any other cases, use the first day of the month.

11.3. Handling of Partial Dates of Concomitant Medication

The algorithm for determining prior and concomitant medications is described below.

A conservative principle will be applied; that is, if there is any missing or partially missing medication start and/or stop date, and it cannot be determined definitively whether a medication was taken concomitantly with study drug, it will be assumed that the medication is concomitant.

If a medication stop date is partially known, the medication stop date will be imputed as latest possible date (ie, last day of month if day is unknown), or 31 December if both the day and month are unknown. If a medication start date is partially known, the medication start date will

be imputed as the earliest possible date (ie, first day of month if day is unknown), or 01 January if both day and month are unknown.

11.4. Reporting Precision

Summary statistics will be presented to the degree of precision listed in Table 10, unless otherwise specified.

Table 10: Degree of Precision

Statistics	Degree of Precision
Mean, Geometric mean, Median, Quartiles, Confidence limit boundaries	One more than the raw data, up to 3 decimal places.
Standard deviation, Standard error	One more than the mean, up to 3 decimal places
Minimum, Maximum	The same as the raw data, up to 2 decimal places
p-value	Rounded to 3 decimal places and therefore presented as 0.xxx; p-values smaller than 0.001 as '<0.001'; p-values greater than 0.999 as '>0.999'.
Percentage	One decimal place. A percentage of 100% will be reported as 100%. Percentages of zero will be reported as 0.

Fractional numeric values will be presented with a zero to the left of the decimal point (for example, 0.12–0.30).

For weight, height, and BMI, one decimal place will be used for summary statistics.

12. REFERENCES

Basen-Engquist K, Bodurka-Bevers D, Fitzgerald MA, et al. Reliability and validity of the functional assessment of cancer therapy-ovarian. *J Clin Oncol.* 2001;19(6):1809-1817. doi: 10.1200/JCO.2001.19.6.1809.

Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. *Biometrics.* 1982;38(1):29-41. doi: 10.2307/2530286.

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics - Theory and Methods.* 1991;20(8):2609-2631. doi: 10.1080/03610929108830654.

van Hout B, Janssen MF, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value Health.* 2012;15(Jul-Aug):708-715. doi: 10.1016/j.jval.2012.02.008.

13. APPENDIX

13.1. Extended Time Without an Evaluable Tumor Assessment

Given the scheduled tumor assessment, ie, q6w (±7 days) during the first year of study participation and q12w (±7 days) thereafter, the following rules will be used to calculate the missed tumor assessment for PFS to identify extended time without an evaluable tumor assessment.

If the duration between the event (PD/death) and last evaluable tumor assessment prior to the new anticancer treatment is more than the window, PFS will be censored at the last evaluable tumor assessment prior to the event (PD/death) and the new anticancer treatment; the date of randomization will be used if there are no post-baseline evaluable assessments.

Define Event Day (since randomization) and Duration (without an evaluable tumor assessment) as below:

- Event Day = min(PD/Death date) − randomization date + 1
- Duration = min(PD/Death date) − last evaluable assessment date + 1 if with post-baseline evaluable assessments

Table 11 presents the situations of PD or death documented ≥2 missed tumor assessments; otherwise, it will be considered as PD or death documented ≤1 missed tumor assessments.

Table 11: Determination of ≥2 Missed Tumor Assessments

Scenario	Considered as ≥2 Missed Tumor Assessments
With last evaluable tumor assessment	91 days < Event Day ≤385 days, and Duration >98 days
	385 days < Event Day <539 days, and Duration >140 days
	Event Day ≥ 539 days, and Duration >182 days
No baseline or no post-baseline evaluable tumor assessment	Event Day >91 days