Statistical Analysis Plan

Clinical Trial Protocol Identification No.	EMR 100070-005
Title:	A Phase III open-label, multicenter trial of avelumab (MSB0010718C) versus platinum-based doublet as a first line treatment of recurrent or Stage IV PD-L1+ non-small cell lung cancer
Trial Phase	Phase III
Investigational Medicinal Product(s)	Avelumab
Clinical Trial Protocol Version	03 January 2019/Version 6.0
Statistical Analysis Plan Author	PPD
Statistical Analysis Plan Date and Version	12 October 2021/Version 6.0
Statistical Analysis Plan Reviewers	PPD , PPD PPD , Senior Expert Statistician PPD , PPD Medical Responsible PPD PPD PPD , PPD PPD , PPD PPD , PPD Biostatistics Program Lead PPD PPD , PPD

Approval Page

Integrated Analysis Plan: EMR 100070-005

A Phase III open-label, multicenter trial of avelumab (MSB0010718C) versus platinum-based doublet as a first line treatment of recurrent or Stage IV PD-L1+ non-small cell lung cancer

Approval of the IAP by all Merck Data Analysis Responsible has to be documented within ELDORADO via eSignature. With the approval, the Merck responsible for each of the analysis also takes responsibility that all reviewers' comments are addressed adequately.

By using eSignature, the signature will appear at the end of the document.

1 Table of Contents

Table of Tables

2 List of Abbreviations and Definition of Terms

ADA	Anti-drug antibody
AE(s)	Adverse event(s)
AESI	Adverse event of special interest
ALK	Anaplastic lymphoma kinase
ALT	Alanine aminotransferase
ANC	Absolute neutrophil count
aPTT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
BOR	Best overall response
BSA	Body Surface Area
Ceoi	Concentration observed immediately at the end of infusion
CI	Confidence interval(s)
CIPD	Clinically important protocol deviations
CMH	Cochran-Mantel-Haenszel
CPI	Checkpoint inhibitors
CPK	Creatine Kinase
CR	Complete response
CRF	Case Report Form
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
Ctrough	Concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration for multiple dosing)
DBP	Diastolic blood pressure
EAIR	Exposure Adjusted Incidence Rate
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCDF	Empirical cumulative distribution function
eCRF	Electronic case report form
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EEA	European Economic Area
EGFR	Epidermal growth factor receptor
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D	EuroQOL five dimensions questionnaire

EuroQOL	European Quality of Life
FAS	Full Analysis Set
FU	Follow-up
GCP	Good Clinical Practice
GGT	Gamma-glutamyl transferase
HB	Hemoglobin
HR	Hazard ratio
HRQoL	Health-Related Quality of Life
ICF	Informed consent form
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
IHC	Immunohistochemistry
IRC	Independent Review Committee
INR	International normalized ratio
IPD	Important Protocol Deviations
irAE	Immune-related adverse event
IRR	Infusion related reaction
IV	Intravenous
ITT	Intention-to-treat
IWRS	Interactive web response system
LDH	Lactate dehydrogenase
LLN	Lower limit of normal
LLOQ	Lower Limit of Quantitation
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
MCIC	Minimal Clinically Important Change
MCV	Mean Corpuscular Volume
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-effect model repeated measures
MRI	Magnetic resonance imaging
nAb	Neutralizing antibody
NCI	National Cancer Institute
NE	Non-evaluable
NSCLC	Non-small cell lung cancer
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PD-1	Programmed death 1

PD-L1	Programmed death ligand 1
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PLT	Platelet count
PR	Partial response
PRO(s)	Patient-Reported Outcome(s)
PT	Preferred term
QLQ-C30	Quality of Life Questionnaire-Core 30 items
QLQ-LC13	Quality of Life Questionnaire-Lung cancer 13-item module
QOL	Quality of Life
RBC	Red blood cell
RECIST 1.1	Response Evaluation Criteria in Solid Tumors version 1.1
RMST	Restricted Mean Survival Time
RPSFT	Rank Preserving Structural Failure Time
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SBP	Systolic blood pressure
SD	Stable disease
SDG	Standardized Drug Grouping
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TMB	Tumor Mutational Burden
TUDD	Time Until Definitive Deterioration
ULOQ	Upper Limit of Quantitation
ULN	Upper limit of normal
VAS	Visual analogue scale
USA	United States

3 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	14 July 2015	PPD	N/A First version
2.0	27 July 2017	PPD	See Section 4.1 Changes to Previous Version
3.0	11 February 2019	PPD	See Section 4.1 Changes to Previous Version
4.0	23 May 2019	PPD	See Section 4.1 Changes to Previous Version
5.0	05Aug2021	PPD	See Section 4.1 Changes to Previous Version
6.0	12Oct2021	PPD	See Section 4.1 Changes to Previous Version

4 Purpose of the Statistical Analysis Plan

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specifications for the interim and final analysis of data collected for protocol EMR100070-005. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). Additionally, the planned analyses identified in this SAP may be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective SAP will be clearly identified in the CSR.

The SAP is based upon Section 8 (Statistics) of the trial protocol and is prepared in compliance with ICH E9. Version 5.0 was created based on the protocol version 6.0 dated 03 January 2019.

4.1 Changes to Previous Version

Version 6.0

The following changes are made in version 6.0 of the SAP:

1. Update on Pharmacokinetics Section 15.3.5.
    a. PK Data Handling section added (15.3.5.1)
    b. Presenting PK concentration data in Listings, Tables and Figures added (15.3.5.2)
2. Update on Population Pharmacometric Analysis Section 15.3.6: reference to Pharmacometric Modeling Analysis Plan regarding analysis and analysis set.
3. ADA categories described in Table 13 in Section 16.5.3 Immunogenicity were updated. The definitions of "Treatment induced", "Treatment-induced Transient positive",

"Treatment-induced Persistent positive", and "Treatment emergent" were added and/or updated.

4. nAb categories described in Table 15 in Section 16.5.3 Immunogenicity were updated. The definitions of "Treatment induced", "Treatment-induced Transient positive", and "Treatment-induced Persistent positive" were updated.

Version 5.0

The following changes are made in version 5.0 of the SAP:

1. Added Section 8.1 to described the summary of the impact of the COVID-19 pandemic on the study and added a reference to potential summaries of COVID-19 vaccinations (Section 13.1)

2. Added subgroups based on tumor mutational burden to Section 8.2.

3. Added description of statistical outputs that will be repeated for Japanese subjects to Section 8.2.

4. Addition of instructions to Section 10 for creation of an average when duplicate central lab measurements exist with the same visit, sample collection date, and sample collection time.

5. Updates to Health-Related Quality of Life analyses (Section 15.2.2).

    a. Changed the denominator for compliance and completion summaries to the number in the FAS, MFAS, or subset based on PD-L1 expression status.

    b. Clarify the descriptive statistics will be provided for visits with ≥ 10 subjects in each treatment group

    c. For MMRM analysis, clarify that measurements after treatment discontinuation can be included.

    d. For TUDD analyses, refer to Minimal Clinically Important Chang (MCIC) rather than Minimum Important Difference (MID) in the definition of definitive deterioration. Also, clarify that deterioration is confirmed by no further improvement of HRQoL score exceeding MCIC magnitude at next assessment or no further available HRQoL data due to death, occurring within 2 scheduled PRO assessments. A table was added to describe event and censoring rules. Finally, remove the sentence that definitive deterioration must be within 5 days after treatment discontinuation to be considered for analysis.

6. Updated description of eDISH plot in Section 16.3.1 to indicate that peak total bilirubin (/ULN) will be displayed rather than the bilirubin (/ULN) result that was concurrent with the peak ALT or AST.

7. Update Table 16 to indicate that an IRR identified as a sign or symptom is required to have should have resolved on the day of onset or the next day rather than within 2 days of onset.

Version 4.0

The following changes are made in version 4.0 of the SAP:

1. Added the modified safety analysis set to Section 8.2 and added specific tables which will be produced for this analysis set in Section 16.

2. Added text to describe imputation rules for laboratory results reported as <LLOQ or >ULOQ to Section 10.

3. Added instructions to Section 10 for selection of records for summaries of lab assessments by visit when multiple assessments are available for a specific visit.

4. Added two further subgroups to Section 8.2.

Version 3.0

The following changes are made in version 3.0 of the SAP:

1. The following sections were updated to align with protocol v6.0. Changes to the protocol included revisons to assumptions on PFS drop-out rate. Also, the scheduling of interim and primary analysis of PFS and OS were changed to have interim and primary analysis of both endpoints at the same time determined by the number of OS events and minimum-follow-up time.

    a. Section 5 Summary of Clinical Trial Features,

    b. Section 6.1 Sample Size,

    c. Section 6.2 Randomization.

    d. Section 7 Overview of Planned Analyses, including all subsections

    e. Section 15 Endpoint Evaluation

2. In various sections of the document, changes were made to indicate that there will be one interim analysis and one primary analysis of OS and PFS.

3. Section 8.2 Analysis Sets:

    a. The modified full analysis set (mFAS) definition was updated to account for the 36 subjects randomized according to the outdated protocol version.

    b. Subgroups defined by the presence or absence of liver metastases, bone metastases, visceral metastases, and pleural effusion at baseline were added.

4. Section 12.3.1 - Stage at initial diagnosis and study entry were added to the list of disease characteristics

5. Section 15.1.2 Progression Free Survival:

    a. Added clarification regarding the interval used to determine when subjects missed two scans into Table 5.

    b. Categorized censoring reasons as administrative and non-administrative

6. Section 15.3.1 Progression-free survival on next-line therapy (PFS2) added

7. Section 15.1.4 Sensitivity Analyses of the Primary Endpoint: The following analyses were added:

    a. PFS where subjects are not censored based on their subsequent anti-cancer treatment.

    b. PFS in treated subjects, i.e. safety analysis set.

8. Section 16.1 text was added to describe the 3-tier approach to summarize and analyze AEs
9. Section 16.2.3 Other Significant Adverse Events was updated to indicate that irAEs are identified based on a list of MedDRA PTs and other qualifying criteria (for details see Appendix V); related irAE were deleted from the irAE analysis.

Version 2.0

The following changes are made in version 2.0 of the SAP:

1. The term "Intention-to-Treat (ITT) analysis set" was updated as "Full analysis set (FAS)" throughout this document (the definition remains the same) per protocol version 4.0;
2. Per-protocol (PP) analysis set and analysis were deleted as no PP analysis will be performed on this study per protocol version 4.0;
3. Modified Full Analysis Set (mFAS) was added for the specific comparison of Arm B vs. Arm C, due to the addition of Arm C in the study and the fact that recruitment of patients in Arm C started after recruitment of patients in Arm A and Arm B;
4. Section 5 Summary of Clinical Trial Features was updated per protocol version 4.0, in particular to present the addition of the third arm (avelumab arm C) and the change in study endpoints (primary endpoints Progression Free Survival – PFS – and Overall survival – OS);
5. Section 6.1 Sample Size was updated for the total number of randomized subjects to be 1095, based on updated calculation per protocol version 4.0;
6. Section 6.2 Randomization was updated to add randomization strata of PD-L1 tumor expression levels at baseline per protocol version 4.0;

7. Section 7 Overview of Planned Analyses was updated to include interim analyses and O'Brien-Fleming boundaries based on a Lan-DeMets alpha spending function that account for the actual number of PFS events in subjects with high PD-L1+ expression tumors at the interim and the final analysis; data cut-off date for the interim and the final analysis was updated as pre-specified date based on a pre-specified expected number of events;
8. Section 8.2 Health-Related Quality of Life (HRQoL) analysis set and modified HRQoL analysis set were added; Subgroup of PD-L1++ status Yes/No was replaced by subgroup PD-L1+ Expression level Low/Moderate/High; Safety Cohort analysis set was added;
9. Section 10 General Specifications for Statistical Analyses was updated; Reference to listings for safety cohort was added;
10. Section 11.1 Protocol Deviations was updated with standard text; clinically important protocol deviations (CIPDs) were added;
11. Section 12.1 Demographics – age groups and region classification were updated;
12. Section 13.1 Prior and Concomitant Medications/Procedures, definition of prior medications was updated;
13. Section 13.2 Subsequent Anti-Cancer Therapies/Procedures was updated with more details;
14. Section 14 Treatment Compliance and Exposure was updated to add calculations for avelumab arm C;
15. Sections 15.1 Primary Endpoint Analyses and 15.2 Secondary Endpoint Analyses were updated to reflect changes of study endpoints per protocol version 4.0;
16. Section 15.1.4 Sensitivity Analyses was updated to include Restricted Mean Survival Time (RMST) as an alternate method to estimate the effect size for time-to-event endpoints in case proportional hazards assumptions cannot be hold. Additional sensitivity analysis was added;
17. Section 15.2.1 Best Overall Response (BOR) was updated to clarify tumor assessment after start of any further anti-cancer will not be included in BOR derivation; Table 6 was updated to cite the original BOR derivation table from Eisenhauser, et al.;
18. Section 15.2.2 Health Related Quality of Life was updated to include more details on the questionnaires scoring system; compliance, time until definitive deterioration (TUDD) analysis, and additional summary analysis including figures were added; details on analysis summary of scheduled and unscheduled visits were added;
19. Section 15.3.5 Descriptive Pharmacokinetic (PK) Analysis was updated based on PK team input;
20. Section 16.1 Adverse Events, definition of immune-related AE (irAE) and Infusion-related reactions (IRRs) was updated; details of pooling the same AE with different toxicity grade, outcome or seriousness recorded as different entries on the eCRF were added;
21. Section 16.2.3 Other Significant Adverse Events was updated to add description of Immune-related adverse event (irAEs) and IRRs;

22. Section 16.2.4 Immunogenicity Subgroup Analysis of Adverse Events was added;
23. Section 16.3.1 Hematology and Chemistry Parameters was updated;
24. Section 16.5.3 Immunogenicity details were added for the immunogenicity data analysis;
25. Appendix II was updated to include important protocol deviations by programming check and medical review;
26. Appendix III Pre-specified Search List of MedDRA Preferred Term for Immune Related Adverse Events was removed from this SAP. A version-controlled search list will be available in Sponsor's MARVEL system;
27. Appendix III was added to include the screenshot of EORTC QLQ-C30 version 3.0 scoring system;
28. Appendix IV was added to include the screenshot of EORTC QLQ-C30-LC13 Lung Cancer Module scoring system;
29. Appendix V was added to include "Description of the Case Review for Assessment of Immune-Related AEs" and detailed information for IRR.

5 Summary of Clinical Trial Features

Trial objectives	**Primary objective** The primary objective is to demonstrate superiority with regard to overall survival (OS) or progression-free survival (PFS) of avelumab versus platinum-based doublet, based on an Independent Review Committee assessment as per RECIST 1.1, in NSCLC subjects with high expression PD-L1+ tumors. **Secondary objectives** Secondary objectives are as follows: • To demonstrate superiority with regard to OS or PFS based on an independent review committee assessment per RECIST 1.1 in NSCLC subjects with moderate and high expression PD-L1+ tumors • To demonstrate superiority with regard to OS in NSCLC subjects with any expression PD-L1+ tumors • To comparatively assess the objective response rate (ORR) by RECIST 1.1 of avelumab versus chemotherapy in high, moderate and high, and any expression PD-L1+ tumors • To determine duration of response of avelumab versus chemotherapy To compare the patient-reported outcomes / quality of life when treated with avelumab versus chemotherapy using the European Quality of Life (EuroQOL) 5-dimensions questionnaire (EQ-5D) and the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 and module QLQ-LC13

	• To determine the safety and tolerability of avelumab **Exploratory objectives** CCI [redacted]
Trial design and plan	**Methodology:** This is a multicenter, international, randomized, open-label, Phase III trial in chemotherapy-naïve (first line) metastatic or recurrent (Stage IV) NSCLC subjects comparing avelumab to first-line platinum-based chemotherapy. The trial consists of a 28-day screening period, followed by the treatment phase (within 4 days after randomization). To achieve the target enrollment of 484 subjects randomized with high expression PD-L1+ tumors, approximately 1131 eligible subjects will be randomized with any expression PD-L1+ tumors and approximately 3100 subjects will be screened for study participation. Subjects will be randomly allocated into the 3 study arms as follows: • Arm A: Avelumab at a dose of 10 mg/kg as a 1-hour (-10/+20 minutes) intravenous (IV) infusion once every 2 weeks until disease progression or unacceptable toxicities, or • Arm B: Investigator's choice platinum containing chemotherapy regimen to be administered in 3-week cycles up to a maximum of 6 cycles of IV

injection until disease progression or unacceptable toxicities consisting of one of the following:

- for patients whose tumor is of non-squamous histology:
  - pemetrexed (500 mg/m^2) in combination with cisplatin (75 mg/m^2 administered on Day 1 of each cycle) or carboplatin (AUC 6 mg/mL x min administered on Day 1 of each cycle).
- for patients whose tumor is of squamous histology:
  - paclitaxel (200 mg/m^2) plus carboplatin (AUC 6 mg/mL x min administered on Day 1 of each cycle); or
  - gemcitabine (1250 mg/m^2 administered on Day 1 and Day 8) plus cisplatin (75 mg/m^2); or
  - gemcitabine (1000 mg/m^2 administered on Day 1 and Day 8) plus carboplatin (AUC 5 mg/mL x min).

- Arm C: Avelumab at a dose of 10 mg/kg as a 1-hour (-10/+20 minutes) IV infusion every week for 12 consecutive weeks, followed by avelumab at a dose of 10 mg/kg once every 2 weeks until disease progression or unacceptable toxicities

Subjects will be randomly allocated to one of the 3 treatment arms, initially into Arm A and Arm B in a 1:1 ratio. The allocation ratio changed to 1:2:2 (Arm A: Arm B: Arm C) once the IDMC recommended, July 2017, the avelumab once a week dosing regimen to be included in the randomization scheme after the initial safety evaluation of avelumab 10 mg/kg every week in a cohort of 6 subjects.

NSCLC histology (squamous versus non-squamous cell) will be used as stratification factors for randomization. The PD-L1 tumor expression level at baseline (low expression versus moderate expression versus high expression) will be supplemented as randomization strata for the 1:2:2 randomization scheme.

Tumor measurements by computed tomography (CT) scan or magnetic resonance imaging (MRI) will be performed every 6 weeks for the first 12 months and every 12 weeks thereafter to determine response to treatment. Response will be evaluated using the Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1).

Treatment with avelumab will continue until disease progression or unacceptable toxicity. Subjects receiving avelumab who have experienced a complete response (CR) should be treated for a minimum of 12 months and/or until disease progression or unacceptable toxicity, after confirmation of response. In case a subject with a confirmed CR relapses after stopping treatment, one re-initiation of treatment is allowed at the discretion of the Investigator and agreement of the medical monitor.

	Treatment with chemotherapy will continue until disease progression or unacceptable toxicity or after the completion of 6 cycles of chemotherapy. Subjects with non-squamous histology are authorized to continue to receive pemetrexed as a maintenance therapy after 4 cycles of platinum-based chemotherapy if their disease has not progressed, or in accordance with the pemetrexed local label. No other maintenance therapy is permitted. Decisions regarding medical management of subjects will be made by the Investigator; however, the primary and secondary endpoint determinations (response and progressive disease [PD]) will be according to the central imaging assessment and review by a blinded Independent Review Committee (IRC). Adverse events (AEs) will be assessed throughout the trial and evaluated using the National Cancer Institute (NCI) Common Technology Criteria version for Adverse Events version 4.03 (CTCAE v 4.03). Periodic evaluations of the trial data will be conducted by an Independent Data Monitoring Committee (IDMC) to ensure subject safety, and the validity and scientific merit of the trial. • **End-of-Treatment visit**: within 7 days from the decision to discontinue, or before the start of any other antineoplastic therapy including a full safety evaluation for subjects that have discontinued treatment due to an AE, and • **Follow-up phase**: Safety Follow-up visit 30 days (± 5 days) after the last administration of trial treatment, followed by a 90-day telephone Safety Follow-up (± 1 week), and Long-term Follow-up every 12 weeks (± 1 week).
Planned number of subjects	Approximately 3100 subjects will be screened. Accrual will proceed up to a target number of approximately 1131 subjects randomized with any expression PD-L1+ tumors in order to obtain at least 484 subjects randomized with high expression PDL1+ tumors. This is based on the estimation that among subjects with any expression PD-L1+ tumors, approximately 45% of them will have high expression PD-L1 tumors.
Primary endpoint	The primary endpoints for the trial are PFS and OS. PFS is defined as the time from date of randomization until date of the first documentation of PD as determined by the IRC (per RECIST 1.1) or death due to any cause in the absence of documented PD, whichever occurs first. OS is defined as the time from randomization to the date of death, regardless of the actual cause of the subject's death.
Secondary/ CCI endpoints	The secondary endpoints include: • BOR according to RECIST 1.1 and as adjudicated by the IRC • Duration of response according to RECIST 1.1 • Patient-reported outcomes/quality of life (assessed by the EQ-5D 5L, and the EORTC QLQ-C30, and module QLQ-LC13 questionnaires)

	• Safety endpoints (including AEs, clinical laboratory assessments, vital signs, physical examination, electrocardiogram [ECG] parameters, and ECOG PS). CCI [redacted]

CCI [redacted]

CCI

[redacted]

6.2 Randomization

Qualified subjects will be randomly allocated to the treatment groups to receive avelumab 10 mg/kg once every 2 weeks (Arm A) or chemotherapy (Arm B) or avelumab 10 mg/kg once a week for 12 consecutive weeks, followed by avelumab at a dose of 10 mg/kg once every 2 weeks (Arm C), using stratified permuted block randomization via the IWRS.

Subjects will be randomly allocated, initially into Arm A and Arm B in a 1:1 ratio.

After the initial safety evaluation of avelumab 10 mg/kg every week in a cohort of 6 subjects was conducted, the IDMC recommended to have the avelumab once a week dosing regimen included in the randomization scheme, changing the allocation ratio to 1:2:2 amongst the three arms (Arm A: Arm B: Arm C). This safety cohort of 10 subjects (4 additional subjects were enrolled in the safety cohort after the initial 6) treated with avelumab 10 mg/kg once a week for 12 consecutive weeks, followed by avelumab at a dose of 10 mg/kg once every 2 weeks was not considered for the randomization process.

The NSCLC histology (squamous versus non-squamous cell) will be used as stratification factor for randomization. The PD-L1 tumor expression level at baseline (low expression versus moderate expression versus high expression) will be supplemented as randomization strata for the 1:2:2 randomization scheme. The purpose of stratification is to ensure balanced distribution of factors with potential prognostic or predictive impact of response factors between treatment arms.

Randomization will occur upon completion of the Screening procedures and determination of subject eligibility, using a stratified permuted block randomization via IWRS as described in Section 6.3 of the study clinical protocol.

7 Overview of Planned Analyses

This SAP covers the analyses for efficacy and safety based on the data cut-off dates for the interim and final analyses. Statistical analyses will be performed using cleaned eCRF data, central laboratory data as well as data of tumor assessment results adjudicated by the Independent Review Committee (IRC) and Health-related quality of life (HRQoL) data and biomarker data, which are collected by an external vendor.

No database can be locked and no randomization code or PD-L1 stratification should be unblinded until this SAP has been approved.

All data will be included up to a prospectively determined clinical cut-off date.

The data cut-off for the interim and the final analysis will be prospectively determined based on event projection provided by the unblinded team. For the interim PFS and OS analysis, the clinical cut-off date will be the date on which approximately 130 OS events for arms C and B are expected in high expression PD-L1+ subjects based on the event projections performed by the unblinded team, and a minimum follow-up of 10 months. For the final PFS and final OS analysis, the clinical cut-off date will be the date on which approximately 173 OS events for arms C and B are expected

in high expression PD-L1+ subjects based on the event projections performed by the unblinded team, and a minimum follow-up of 20 months.

Since the observed number of events at the interim analysis may not be exactly equal to the planned 130 OS events in high expression PD-L1+ subjects, the efficacy boundary will be updated based on the actual number of observed events using the pre-specified alpha-spending function. The observed Z-test statistic at the interim analysis will be compared with the updated efficacy boundary. For the final analysis, if the number of OS events deviates from the target number of 173 OS events in high expression PD-L1+ subjects, the final analysis criteria will be determined, taking into account the actual alpha spent at the interim analysis and the actual association between the two test z-scales as specified in Table 2, Section 7.2 so that the overall one-sided significance level is controlled at 0.025.

Since the formal efficacy boundaries will be used at the interim analysis for the statistical testing of PFS or OS, a statistically significant finding at the interim will be intended to claim superiority.

The planned analyses include:

- An interim PFS and OS analysis
- A final PFS and OS analysis
- A follow-up analysis

A separate SAP covers the periodic safety review and interim PFS and OS analysis by the Independent Data Monitoring Committee (IDMC).

Separate or supplemental analysis plans might be written to cover:

- The analysis for PopPK, exposure/efficacy and exposure/safety data to support the Summary of Clinical Pharmacology when appropriate
- Additional analyses of patient reported outcomes (PRO) (i.e. EQ-5D, EORTC QLQ-C30 and EORTC QLQ-LC13)

CCI [redacted]

7.1 Sequence of Analyses

The following analyses will be performed during this trial. As the data cut-off date for both the interim efficacy and primary analyses will be prospectively determined based on event projection provided by the unblinded team, the actual number of OS events may slightly differ from the planned number as indicated below.

- Interim PFS and OS analysis:
    - This analysis will be performed after approximately 130 OS events (75%) have been observed in high expression PD-L1+ subjects and a minimum follow-up of 10 months in arms B and C (mFAS).
- Final PFS and OS analysis:

- This analysis will be performed after approximately 173 OS events (death) have been observed in high expression PD-L1+ subjects and a minimum follow-up of 20 months in arms B and C (mFAS).

There will be a partial database lock for both the interim and primary analyses.

There will be ongoing periodic safety reviews by the IDMC. Details are provided in the IDMC charter and the IDMC SAP.

7.2 Interim Analysis

This interim analysis will include the analysis of the PFS and OS primary endpoints, secondary endpoints including BOR, and safety. Descriptive statistics will also be reported including subject disposition, demographics, medical history and other baseline characteristics, along with safety information (adverse events, deaths, and clinical laboratory evaluation).

An IDMC will be formed and will be responsible for periodic safety evaluations of the trial as well as the evaluation of this interim efficacy analysis. The IDMC will consist of a group of five experts who will be neither participants in the trial nor employees of the Sponsor of this trial, nor the independent statistical provider who is not a voting member of the IDMC. An IDMC charter provides details about the conduct of the IDMC meeting and decision making rules.

The interim analysis will test the hypotheses for PFS: H_{0_1A}, H_{0_1C}, H_{0_3A}, H_{0_3C} and the hypotheses for OS: H_{0_2A}, H_{0_2C}, H_{0_4A}, H_{0_4C}, H_{0_6A}, H_{0_6C} after the required number of OS events for the interim analysis for Arm C+B have been reached and 10 months have elapsed from the last subject randomized for the high expression PD-L1 analysis set in mFAS (see Table 2).

Primary PFS and OS analyses will be conducted after the required number of OS events for Arm C+B have been reached and at least 20 months have been elapsed from last subject randomized for the high expression PD-L1 analysis set in mFAS. The current plan considers

- 130 OS events for Arm C and B and a minimum follow-up of 10 months to determine the cut-off date for the interim PFS and OS analyses.
- 173 OS events and a minimum follow-up of 20 months for the final PFS and OS analyses.

Table 2 describes the expected number of events, information fraction, the associated cumulative alpha spending and the corresponding expected efficacy boundaries, under proportional hazard assumptions. The actual numbers in particular for the Arm A and B could vary as the event size at the planned cut-off dates is unknown. The alpha levels are however fixed according to Lan-DeMets α-spending with O'Brien-Fleming-like boundaries.

Table 2 Interim Analyses

Primary Inferential Analyses (high expression PD-L1)	Expected number of events	Information fraction	Cumulative alpha spent	Efficacy boundary HR
IA – PFS				
Arm A vs. B (FAS)	228	91%	0.0026994	0.69
Arm C vs. B (mFAS)	146	84%	0.0063969	0.66
Final PFS				
Arm A vs. B (FAS)	250	100%	0.004170	0.71
Arm C vs. B (mFAS)	174	100%	0.012500	0.70
IA – OS				
Arm A vs. B (FAS)	214	83%	0.0007241	0.64
Arm C vs. B (mFAS)	130	75%	0.001609	0.60
Final OS				
Arm A vs. B (FAS)	258	100%	0.002080	0.69
Arm C vs. B (mFAS)	173	100%	0.006250	0.68

FAS: full analysis set, HR: hazard ratio, IA: interim analysis, mFAS: modified full analysis set, OS: overall survival, PFS: progression free survival.

The interim OS and PFS analysis will be conducted on the mFAS for arm C vs. B comparison and on the FAS for arm A vs. B comparison, with the primary analysis population being the high PD-L1+ subset of the mFAS and FAS. The IDMC will also be presented with subject disposition, subject background, baseline disease and demographic information, along with safety information. Details of the IDMC mission, composition, and operations are provided in the IDMC charter.

An independent statistical provider will perform the unblinded interim safety and efficacy analysis to support the IDMC. After the prospectively determined data cut-off date is reached when the determined number of OS events in high expression PD-L1+ subjects is approximately observed for the interim analysis, the independent statistical provider will prepare the outputs (using programs prepared by the blinded team based on dummy treatment and PD-L1+ variables) in agreement with the IDMC charter and transmit the analyses, tabulations, and listings to the IDMC for the meeting. Results from the interim safety and efficacy analysis will be transmitted from the independent, unblinded statistician to the IDMC only. The independent statistician will be available at the IDMC meeting should any questions from the IDMC members arise regarding the data and / or analyses.

A closed testing procedure with hierarchically ordered hypotheses is applied in the study. If the one-sided p-value from the stratified log-rank test of PFS or OS in high expression PD-L1+ subjects is below the significance level at the interim analysis based on a Lan-DeMets alpha spending function for O'Brien-Fleming boundaries (see Table 2), the treatment difference will be claimed as statistically significant and the next step in the hierarchical testing strategy will be tested (see Section 15.1.1). If, at the time of this interim analysis, the PFS or OS of avelumab

assigned subjects is shown to be superior to that of those randomized to the chemotherapy group in high expression PD-L1+ subjects with a p-value below the significance level at the interim analysis based on a Lan-DeMets alpha spending function for O'Brien-Fleming boundaries (see Table 2), the IDMC may declare superior efficacy in the avelumab treatment arm compared with those randomized to receive chemotherapy. This interim analysis of PFS and OS will be performed in a manner identical to the final efficacy analysis. Details of hierarchical testing strategy are provided in Section 15.1.1. In case of a positive interim analysis (i.e. at least one rejected null hypothesis in either the comparison of C versus B or A versus B for PFS or OS) all analyses described in this SAP will be performed to facilitate report writing.

A partial database lock will be performed for the interim analysis. A data review meeting will be held prior to the interim analysis.

7.3 Final Analysis

This analysis will include the final analysis, or so called primary analysis, of the PFS and OS primary endpoints, and analysis of secondary endpoints including BOR, and safety. Descriptive statistics will also be reported including subject disposition, demographics, medical history, and other baseline characteristics, along with safety information (adverse events, deaths, and clinical laboratory evaluation).

For this trial the final efficacy analysis will be conducted after the prospectively determined data cut-off date when the number of OS events in high expression PD-L1+ subjects reaches approximately 173 in arms B and C (mFAS) based on the event projection provided by the unblinded team and a minimum follow-up of 20 months is observed. The O'Brien-Fleming information fraction for type I error and the efficacy boundaries based on a Lan-DeMets spending function are presented in Table 2 for different number of PFS and OS events in high expression PD-L1+ subjects at the interim analysis and will be adjusted for final PFS and OS analysis. No inferential efficacy analysis will be performed for hypotheses that are rejected at the interim analysis. In this situation, analysis may be conducted to enhance precision of estimates after the study is determined to stop early at the interim analysis. Hypotheses that cannot be rejected at interim can be tested at the final analysis following the hierarchical testing procedure at the remaining significance level based on the Lan-DeMets alpha spending function for O'Brien-Fleming boundaries to control the overall Type 1 error rate at 0.025 (one-sided).

Study staff involved with the day to day management of the trial, as well as any sponsor staff, will not have access to the results unblinded to both treatment group and expression PD-L1+ status unless the trial is stopped at interim as recommended by the IDMC.

A partial database lock will be performed for the final PFS and OS analysis. A data review meeting will be held prior to the final PFS and OS analysis.

7.4 Follow-up analysis

Subject follow-up for progression and survival will continue until 5 years after the last subject receives the last dose of avelumab. The Sponsor may terminate the study at any time once access

to investigational medicinal product (IMP) for subjects still benefitting is provisioned via a roll over study, expanded access, marketed product or another mechanism of access as appropriate.

Therefore, the full database lock will take place either 5 years after the last subject receives the last dose of avelumab or after the Sponsor decides to terminate the study, whichever occurs first.

Follow-up analysis will include subject status and survival follow-up information in a listing. Depending on the status of the study, additional variables, e.g. for safety, may also be included.

8 Changes to the Planned Analyses in the Clinical Trial Protocol

The statistical methods specified in this document are in accordance with protocol version 6.0 (dated 03 January 2019).

8.1 COVID-19 Impact

No changes to the planned analysis of the efficacy endpoints will be performed due to the impact of Coronavirus disease 2019 (COVID-19) outbreak.

Additional outputs (summary table and listing) will be generated to describe the impact of COVID-19 in terms of the following:

- On-treatment during the pandemic
- On-study during the pandemic
- Potentially affected by COVID-19
- Adverse events
- COVID-19 vaccinations
- COVID-19 vaccination associated AEs
- Protocol deviations (important and non-important)
- Missed visits (including number of missed visits)
- Missed or delayed efficacy evaluations
- Drug administration - missed doses
- Treatment discontinuation
- Discontinuation of study sssessments
- Study discontinuation
- Death

The start of COVID-19 pandemic will be defined by country as the earlier of the date of the first death from COVID-19 according to the published data by European Centre for Disease Prevention and Control on 26th June 2020 (https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide) or 11th March 2020 (when the WHO declared the COVID-19 pandemic).

COVID-19 related AEs defined by the MedDRA COVID-19 SMQ for the specific MedDRA version used for final coding. Summaries will include:

- A table with TEAEs associated to COVID-19, when ≥ 10 patients with AEs are observed.
- A listing comprising all COVID-19 related AE terms.

The WHO-DD standardized drug grouping (SDG) "Vaccines for COVID-19" is a subgroup of the overarching SDG "Drugs and vaccines for COVID-19". The vaccine-related SDG is further subdivided into subgroup SDGs identifying type of vaccine (e.g. DNA vaccines for COVID-19, Inactivated vaccines for COVID-19, etc.). A table with overall number of patients with vaccinations received by SDG subgroups and product name (preferred term) will be produced if ≥ 10 patients with vaccinations are observed. A listing will be created regardless of the number of subjects vaccinated.

A frequency table will be produced for the full analysis set to present the number of subjects with any protocol deviations related to COVID-19 (categorized by frequency of subjects with a protocol deviation overall as well as by category and type of protocol deviation). Important and non-important protocols deviations will be combined for this summary. In addition, a listing of protocol deviations attributed to COVID-19 will also be produced.

8.2 Disposition of Subjects and Discontinuations

Analysis sets: Screening Analysis Set / FAS / mFAS by PD-L1 expression status / Safety Cohort Analysis Set

The following will be summarized overall and per treatment group in the screening analysis set. The summary will also be presented for all FAS subjects and mFAS subjects, by expression PD-L1+ status (any, moderate-and-high, high), starting from "Number and percentage of randomized subjects". The percentages below will be calculated based on the number of subjects in the FAS and mFAS by expression PD-L1+ status, respectively:

- Total number of subjects screened overall
- Number of subjects who discontinued from the trial prior to randomization overall and grouped by the main reason (i.e. screen failures)
- Number and percentage of randomized subjects in the following populations:
  - FAS
  - mFAS
  - HRQoL-AS
  - mHRQoL-AS
  - Safety-AS
  - PK analysis set
- Number and percentage of subjects randomized but not treated
- Number and percentage of subjects who received at least one dose of treatment
- Number and percentage of randomized subjects still on treatment

- Number and percentage of randomized subjects who completed treatment

  Note: treatment completed if death, PD or 6-cycles of chemotherapy for at least one chemotherapy completed

- Number and percentage of randomized subjects who discontinued treatment and reasons for treatment discontinuation
- Number and percentage of randomized subjects who discontinued one drug in chemotherapy arm but are still on treatment
- Number and percentage of subjects who discontinued or completed the treatment but are still in follow-up including number of subjects still in follow-up for PFS and those still in follow-up for OS
- Number and percentage of subjects who re-initiated avelumab treatment, and number of completions (death or PD) and discontinuations after re-initiation
- Number and percentage of subjects who switched to avelumab after progression in the chemotherapy arm, and number of completions (death or PD) and discontinuations and discontinuation reasons after switch
- Number and percentage of randomized subjects who discontinued trial
- Number and percentage of randomized subjects per reason for trial discontinuation

The results of the randomization algorithm (according to IWRS) will be summarized as follows for all randomized subjects:

- Number and percentage of randomized subjects overall, by region (European Economic Area (EEA), required by EudraCT), North America, Latin America, Asia, Australasia, Africa and Middle East), by country within region
- Number and percentage of randomized subjects by center
- Number and percentage of randomized subjects by randomization strata (IWRS)
- Number and percentage of randomized subjects by randomization strata (eCRF)
- Cross tabulation: stratum by IWRS vs. stratum by eCRF (histology only)
- Cross tabulation: subjects randomized (avelumab/chemotherapy/avelumab weekly) vs. subjects treated (avelumab/chemotherapy/avelumab weekly/not treated)
    - Reasons if randomized is different from treated

9 Analysis Sets

Screening Analysis Set

The screening analysis set will include all subjects who signed the informed consent.

Full Analysis Set

The full analysis set (FAS) will include all subjects who were randomized to study treatment. Analyses performed on the FAS will take into account subjects' allocation to treatment groups as randomized. For subjects who are randomized more than once with different subject identifier, the first randomization will be used in the analysis set.

Modified Full Analysis Set

The modified FAS (mFAS) is defined as the FAS restricted to subjects who were randomly assigned to treatment according to CTP version 4.0 or later, after Arm C was activated in the treatment allocation.

The 36 subjects who were randomized after implementation of protocol version 4.0, but according to protocol version 3.0, will be included in the FAS but not in the mFAS.

HRQoL Analysis Set

The HRQoL analysis set (HRQoL-AS) is a subset of the FAS and will include all FAS subjects who meet both of the following criteria:

- Have at least one baseline HRQoL questionnaire completed
- Have at least one post-baseline HRQoL questionnaire completed

Modified HRQoL Analysis Set

The modified HRQoL analysis set (mHRQoL AS) is a subset of the mFAS and will include all mFAS subjects who meet both of the following criteria:

- Have at least one baseline HRQoL questionnaire completed
- Have at least one post-baseline HRQoL questionnaire completed

Safety Analysis Set

The Safety analysis set (Safety-AS) will include all randomized subjects who were administered at least one dose of the study medication, i.e. avelumab or chemotherapy. Analyses performed on the Safety analysis set will take into account subjects' allocation to treatment groups as treated: subjects will be classified according to the treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case patients will be classified according to the first study treatment received.

Modified Safety Analysis Set

The modified Safety-AS (mSafety-AS) is defined as the Safety-AS restricted to subjects who were randomly assigned to treatment according to CTP version 4.0 or later, after Arm C was activated in the treatment allocation.

Safety Cohort Analysis Set

The Safety Cohort analysis set will include all subjects enrolled (not randomized) in the safety cohort who have received at least one dose of avelumab. For these subjects, data listings will be provided, and a "Safety Cohort" column will be included on selected safety tables.

PK Analysis Set

PK Analysis set is a subset of the safety analysis set and will include all subjects in the avelumab treatment arms who have at least one measurable post-dose drug concentration.

PD-L1 Expression Positivity Criteria

The inclusion criteria for positive PD-L1 is determined by a Dako PD-L1 immunohistochemistry (IHC) pharmDx companion diagnostic test, validated for the purpose of this trial. The PD-L1 IHC assay and the scoring algorithm to determine low, moderate, and high PD-L1 expression level will be defined and validated prior to conducting any statistical analyses. Details are specified below:

- High expression PD-L1+: Tumor cells deemed positive for PD-L1 expression (PD-L1+) at any staining intensity (1+, 2+ or 3+) in ≥80% and ≤100% of tumor cells
- Moderate expression PD-L1+: Tumor cells deemed positive for PD-L1 expression (PD-L1+) at any staining intensity (1+, 2+ or 3+) in ≥50% and <80% of tumor cells
- Low expression PD-L1+: Tumor cells deemed positive for PD-L1 expression (PD-L1+) at any staining intensity (1+, 2+ or 3+) in ≥1% and <50% of tumor cells

For purpose of analyses, the following categorization will be considered (cited as PD-L1 status in the following):

- Any expression PD-L1+ tumor cells at any staining intensity (i.e. low, moderate and high)
- Moderate and high expression PD-L1+ tumor cells at any staining intensity (i.e. moderate and high)
- High expression PD-L1+ tumor cells at any staining intensity (i.e. high only)

From protocol version 4, categorization will be captured by the testing lab via the IWRS. PD-L1+ expression classification for subjects enrolled prior to version 4 will be derived from the results of the Dako PD-L1 IHC pharmDx companion diagnostic test. The study team is blinded to PD-L1+ expression.

All analysis sets to be considered are presented in Table 3.

Table 3 Analysis Sets by PD-L1 status

	Screening Analysis Set	Full Analysis Set (FAS)	Modified Analysis Set (mFAS)	HRQoL Analysis Set (HRQoL-AS)	Modified HRQoL Analysis Set (mHRQoL-AS)	Safety Analysis Set (Safety-AS)	PK Analysis Set
Any expression PD-L1+	Screening Analysis Set	FAS	mFAS	HRQoL-AS	mHRQoL-AS	Safety-AS	PK Analysis Set
Moderate and high expression PD-L1+		FAS, mod and high PD-L1+	mFAS, mod and high PD-L1+	HRQoL-AS, mod and high PD-L1+	mHRQoL-AS, mod and high PD-L1+		
High expression PD-L1+		FAS, high PD-L1+	mFAS, high PD-L1+	HRQoL-AS, high PD-L1+	mHRQoL-AS, high PD-L1+		

Table 4 summarizes the use of the analysis sets in the different analyses.

Table 4 Statistical Analysis by Analysis Set

Analyses	FAS	mFAS	HRQoL-AS	mHRQoL-AS	Safety-AS	mSafety-AS	PK Analysis Set
Baseline Characteristics	✓	✓			✓	✓	
Past and Concomitant Therapies	✓	✓					
Important Protocol Deviations	✓	✓					
Compliance and Exposure					✓	✓	
Efficacy: Primary	✓[1]	✓[2]					
Efficacy: Secondary - BOR	✓[1]	✓[2]					
Efficacy: Secondary - HRQoL			✓[1]	✓[2]			
CCI							
Safety					✓	✓[3]	
CCI							

For each analysis set except Safety-AS, any expression PD-L1+, moderate and high expression PD-L1+, and high expression PD-L1+ will be presented separately.

[1] FAS and HRQoL Analysis Set will be used for arm A vs. arm B comparison

[2] mFAS and modified HRQoL Analysis Set will be used for arm C vs. arm B comparison

[3] Selected Safety analyses only

Subgroup Analysis Sets

Subgroup analyses will be performed on primary and secondary efficacy endpoints based on the subgroups as defined below.

For including baseline variables into Cox's proportional hazards model for OS and PFS multivariate analysis, and logistic regression model for BOR multivariate analysis, the following parameterization is to be used. For variables with more than two categories, an indicator variable will be defined for each category except for the first category, which always defines the reference. The final parameterization will be updated and fixed at the Data Review Meeting, at the latest, and documented in an amendment to this SAP if different from the previous definition.

The PD-L1+ status (any, moderate and high, high expression) is part of the hierarchical testing procedure described below in Section 15. The prognostic value of PD-L1+ status is not firmly established and will be analyzed here. It should also be explored if high expression PD-L1+ subjects have a greater treatment effect, i.e., if this is a predictive factor. PD-L1+ level subgroups (low, moderate, high expression) are only relevant to the analyses performed on the FAS and mFAS.

CCI [redacted]. The purpose of subgroups age, gender, region, race, and ethnicity are to check internal consistency within the targeted population. Subgroups histology, ECOG PS and smoking habits are known prognostic factors for lung cancer patients. In case of a low number of subjects within a category (<25 subjects with high PD-L1+ expression from all three treatment groups combined), the categories may be pooled when meaningful. The subgroup analysis will not be performed on any subgroup category "Missing".

The following subgroups will be defined:

- Expression PD-L1+ level
  - Low (Reference)
  - Moderate
  - High
- Age group 1
  - Age <65 years (Reference)
  - Age ≥65 years
- Age Group 2
  - Age < 75 years (Reference)
  - Age ≥ 75 years

- Gender
  - Male (Reference)
  - Female
- Race
  - Caucasian/White (Reference)
  - Asian
  - Black/African American
  - Other or unknown (excluding "missing")
- Ethnicity
  - Hispanic/Latino (Reference)
  - Non-Hispanic/Latino
- Ethnicity
  - Japanese living in Japan (Reference)
  - All other subjects
- Pooled Region
  - US and Western Europe (Reference)
  - Eastern Europe
  - Asia
  - Rest of the World (Australasia, Latin America, Africa and/or Middle East will be included as additional subgroups if including > 10% of the overall randomized/treated population)
- ECOG PS at baseline
  - ECOG PS 0 (Reference)
  - ECOG PS 1
- NSCLC histology as collected on the "Disease History" eCRF page
  - Squamous cell (Reference)
  - Non-squamous cell
- Smoking status
  - Never smoker (Reference)
  - Ever smoker
- Liver metastasis at baseline as collected in the "Site" variable on the target and/or non-target lesions eCRF pages where the "Type" is not NODE

- Absent (Reference)
- Present

- Bone metastasis at baseline as collected in the "Site" variable on the target and/or non-target lesions eCRF pages
  - Absent (Reference)
  - Present
- Visceral metastasis at baseline as collected in the "Site" variable on the target and/or non-target lesions eCRF pages as any from the following list where the "Type" is not NODE: "Oesophagus, Stomach, Small intestines, Pancreas, Liver, Spleen, Kidney, Bladder, Adrenal, Omentum, Colon"
  - Absent (Reference)
  - Present
- Pleural effusion at baseline, collected as "Pleura" in the "Site" variable on the non-target lesions eCRF pages
  - Absent (Reference)
  - Present
- Brain metastasis at baseline as collected on the brain evaluation page or in the "Site" variable on the Prior anti-cancer radio-therapy/surgery eCRF pages
  - Absent (Reference)
  - Present
- Prior Radiotherapy for Thorax as collected in the "Site" variable on the Prior anti-cancer radio-therapy eCRF as follows: Esophagus, Larynx, Lung, Pleura, Mediastinal Nodes, Axillary Nodes, Supraclavicular Nodes, Para-aortic nodes, Breast, Thyroid, Pericardium, spinal cord; or, in the "Site = other" variable as follows (identified by medical review): lung, mediastinum, mediastinal, hilum, chest, lymph node station 7, subcarinal lymphnode:
  - Absent (Reference)
  - Present
- Tumor Mutational Burden (TMB) with tumor only bioinformatics pipeline – the total number of Non-synonymous somatic variants per megabase. Four subgroups will be formed using quantiles:
  - ≤ 25% quantile (Reference)
  - >25% quantile - ≤ 50% quantile
  - >50% quantile - ≤ 75% quantile
  - >75% quantile

Analyses in Japanese subjects

In order to satisfy requirements of the Japanese Pharmaceuticals and Medical Devices Agency, a subset of data summaries will be produced for Japanese subjects. These will include summaries of:

- Subject disposition
- Subjects randomized vs treated
- Analysis sets
- Demographic characteristics
- Disease history
- PD-L1 expression status
- Subsequent anti-cancer drug treatments
- Subsequent immune checkpoint inhibitors
- Avelumab and chemotherapy treatment exposure
- Adverse events (overall, related, grade ≥ 3, related grade ≥ 3, serious)
- Overall survival, progression-free survival, confirmed best overall response, duration of response, CCI [redacted]

10 General Specifications for Statistical Analyses

Unless otherwise indicated all analyses will be presented separately for the three treatment groups.

Baseline characteristics summary and the efficacy analysis will be performed on the FAS and the mFAS separately for any expression PD-L1+, moderate and high expression PD-L1+ and high expression PD-L1+. Analyses performed on the FAS and mFAS will take into account subjects' allocation to treatment groups as randomized. Analyses performed on the safety population will consider subjects as treated.

Derived variables below for the safety cohort will be based on the date of first dose of study drug instead of randomization, whenever applicable.

Data handling after cut-off date:

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

Pooling of centers:

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

Unscheduled visits:

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

Significance level:

The overall significance level is 2.5% one-sided. The confirmatory statistical test for the primary and secondary efficacy endpoint analysis is described in Section 15.1 along with procedures for controlling the overall type I error rate. All other statistical analyses performed on the secondary and other endpoints defined in this SAP are to be regarded as CCI The statistical tests performed on the primary and secondary efficacy endpoints in comparing treatment arms will be one-sided, and the statistical tests to compare treatment arms on other CCI and safety analyses will be two-sided.

Confidence intervals will be two-sided with a confidence level of 95%, if not otherwise specified.

Presentation of continuous and qualitative variables:

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

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

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

No summary statistics or boxplots will be presented if less than 5 subjects are available in at least one treatment group.

Definition of baseline:

For efficacy analyses (including HRQoL):

The last measurement prior to randomization will serve as the baseline measurement. If such a value is missing, the last measurement prior to the first study drug administration will be used as the baseline measurement for the efficacy analysis except for analyses of tumor assessment data where the baseline assessment would be considered missing.

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

For safety analyses:

The last measurement prior to the first study drug administration will be used as the baseline measurement.

If an assessment is planned to be performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first administration of study treatment, it will be assumed that it was performed prior to study treatment administration, if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 for baseline calculation.

Unscheduled assessments will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

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

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

Baseline for HR and QT/QTc assessments will be derived from the visit where both HR and QT are not missing. Unscheduled assessments will not be included in the calculation of the average. QTc assessments will only be derived if not consistently collected for all subjects in the eCRF. In such cases QTc assessments will be derived based on HR and QT.

Even if both central and local labs are collected in the study, the baseline will be derived based only on the central lab collected data.

Definition of duration:

Duration will be calculated by the difference of start and stop date + 1, if not otherwise specified. For example, survival time (days) = date of death – date of randomization + 1.

The time since an event (e.g. time since first diagnosis) will be calculated as reference date minus date of event.

Definition of study day/treatment day:

Study day / Treatment day are defined relative to the date of randomization / start of treatment. Study day 1 defines the day of randomization, the day before is defined as Study day –1 (no Study day 0 is defined). Treatment day will be calculated accordingly, treatment day 1 is defined as the date of first administration of treatment (avelumab or chemotherapy).

Definition of on-treatment period:

On-treatment period is defined as the time from the first study drug administration to the last drug administration date + 30 days or the earliest date of subsequent anti-cancer drug therapy minus 1 day, whichever occurs first, unless otherwise stated.

For subjects randomized to the chemotherapy arm who switch to avelumab after progression, avelumab is also regarded as subsequent anti-cancer drug therapy in the above definition of on-treatment period.

For subjects switching to avelumab, a second on-treatment period is defined similarly as above as the time from the first avelumab administration to the last avelumab administration date + 30 days or earliest date of subsequent anti-cancer drug therapy after the switch to avelumab minus 1 day, whichever occurs first, unless otherwise stated.

Summary statistics over time:

For descriptive statistics over time by nominal visit or time point for HRQoL data and safety endpoints (laboratory, ECG and vital signs), only those visits/time points that have at least 5 subjects in both treatment arms will be included in the summary tables and figures. The exception

is Discontinuation and End-of-Treatment visit which will be included in the summary statistics despite the number of subjects who completed such visit.

Standard derivations and reporting conventions:

The following conversion factors will be used to convert days into weeks, months or years:

1 week = 7 days; 1 month = 30.4375 days, 1 year = 365.25 days.

Demographics and physical measurements:

- Age [years]:
  - (date of given informed consent - date of birth + 1) / 365.25
  - In case of missing day only: Age [years]: (year/month of given informed consent – year/month of birth) / 12
  - In case only year of birth is given: Age [years]: (year of given informed consent - year of birth)

  The integer part of the calculated age will be used for reporting purpose.
- Site codes will be used for the determination of the subject's geographic region.
- BMI (kg/m^2) = weight (kg)/[height (m)]^2.
- BSA (m^2) = ([height (cm) x weight (kg)] / 3600)$^{0.5}$

Smoking history computation:

- Note: 'chewing tobacco' is not taken into account for analysis of smoking history
- Cigarette equivalents are calculated as follows: one cigar is regarded equivalent to 5 cigarettes and 1 pipe is regarded equivalent to 3 cigarettes
- Duration of smoking [years]:

  (end date of smoking – start date of smoking + 1) / 365.25
- Pack-years:
  - calculate cigarette equivalents per day using the conversion factors given above
  - convert to packs per day where 20 cigarettes are regarded as 1 pack
  - pack-years = packs per day × duration of smoking [years]

Disease characteristics computation:

- Time since initial cancer diagnosis (months) =

  (date of randomization – date of initial cancer diagnosis) / 30.4375

- Time since first diagnosis of documented, locally advanced, inoperable or metastatic disease (months) =

  (date of randomization – date of documented, locally advanced, inoperable or metastatic disease) / 30.4375

Prior-anti cancer therapy computation (drug, radiotherapy, surgery):

- Time since last anti-cancer drug therapy (months) =

  (date of randomization – date of last anti-cancer therapy prior randomization) / 30.4375

PD-L1 Expression Level computation:

- Time from sampling date to date of randomization (months) =

  (date of randomization –date of tissue sampling)/30.4375

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

Data collected after switching to avelumab and re-initiated treatment:

Switch refers to subjects who cross over from chemotherapy to avelumab after disease progression confirmed by IRC (option allowed in protocol version 3.0 but no longer in protocol version 4.0). Re-initiation refers to subjects with a confirmed CR who experience relapses after stopping treatment with avelumab and who re-start avelumab treatment.

Data collected after switching to avelumab and re-initiation of treatment will not be included in the main analyses for safety and efficacy except for overall survival , summary of death and disposition as well as for PFS sensitivity analyses as specified below. Separate tables and listings will be created to specifically describe data collected after switch. Data listings will be created for all the data collected after re-initiation of study treatment with a flag variable to indicate such data is collected after the re-initiation of study treatment.

Data collected in safety cohort:

Data collected in the safety cohort will be included in all listings, including those created for randomized subjects. The data for the safety cohort will be presented as a separate treatment arm in these listings as well as in selected safety tables.

Handling of missing data:

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

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

Missing statistics, e.g. when they cannot be calculated, should be presented as "nd". For example, if n=1, the measure of variability (standard deviation) cannot be computed and should be presented as "nd".

Handling of incomplete dates:

Incomplete **dates for disease history** (initial diagnosis date, date of documented, locally advanced, inoperable or metastatic disease diagnosis) will be imputed as follows:

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

For calculation of durations, incomplete **dates for nicotine consumption** (start/end dates) will be imputed as follows:

- As day is not collected, it will be imputed to the 15th day of the month.
- If month is missing and the year is prior to the year of the first study treatment, the month and day will be imputed as July 1st.
- If month is missing and the year is same as the year of the first study treatment, the month and day will be imputed as January 1st.
- If start date is non-missing but end date is missing (current smoker), end date will be imputed by cut-off date.
- If the date of first study treatment is missing, randomization date will be used instead.

Incomplete **AE-related dates** will be imputed as follows:

- If the AE onset date is missing completely, then the onset date will be replaced by the start of study treatment.
- If only the day part of the AE onset date is missing, but the month and year are equal to the start of study treatment, then the AE onset date will be replaced by the start of study treatment. For example, if the AE onset date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed AE onset date will be 15/JAN/2015.

- If both the day and month of the AE onset date are missing but the onset year is equal to the start of study treatment, then the onset date will be replaced by the start of study treatment. For example, if AE onset date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed AE onset date will be 19/NOV/2014.
- In all other cases the missing onset day or missing onset month will be replaced by 1. For example:
    - if AE onset date is --/FEB/2015 and study treatment start date is 03/MAR/2015, then the imputed AE onset date will be 01/FEB/2015.
    - if AE onset date is --/---/2015 and study treatment start date is 12/DEC/2014, then the imputed AE onset date will be 01/JAN/2015.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of subject's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete stop date will not be imputed.

Further information after cut-off (like fatal outcome) might be taken from Safety database and included separately into CSR.

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

- If the medication start date is missing completely, then the medication date will be replaced by the start of study treatment.
- If the day of medication start date is missing, but the month and year are equal to the start of study treatment, then the medication start date will be replaced by the start of study treatment. For example, if the medication start date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed medication start date will be 15/JAN/2015.
- If both the day and month of medication start date are missing but the start year is equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. For example, if the medication start date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed medication start date will be 19/NOV/2014.
- In all other cases the missing medication start day or missing medication start month will be replaced by 1. For example:
    - if medication start date is --/FEB/2015 and study treatment start date is 03/MAR/2015, then the imputed medication start date will be 01/FEB/2015.
    - if medication start date is --/---/2015 and study treatment start date is 12/DEC/2014, then the imputed medication start date will be 01/JAN/2015.
- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of subject's death. In the latter case the date of death will be used to impute the incomplete stop date.
- In all other cases the incomplete medication stop date will not be imputed.

Incomplete **subsequent anti-cancer therapy start dates** will be imputed as follows:

- If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anti-cancer therapy is before that date. In that case, the incomplete anti-cancer therapy start date will be imputed as the end date of the anti-cancer therapy.
- If both day and month are missing, no imputation will be performed.

Incomplete **exposure dates** will be imputed as follows:

- If the study medication start date is missing, it is assumed that the first dose of study treatment medication is given at the randomization date. The randomization date will replace incomplete dates of the first dose of study treatment.
- In case the last date of study drug is incomplete the date of last administration of study drug will be taken from the treatment termination eCRF pages.

Date of last dose of study drug if unknown or partially unknown will be imputed as follows:

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

  = 31DECYYYY, if only Year is available and Year < Year of min (EOT date, death date)
  = Last day of the month, if both Year and Month are available and Year = Year of min (EOT date, death date) and Month < the month of min (EOT date, death date)

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

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

- All patient assessment dates (blood draws (laboratory, PK), vital signs, performance status, ECG, tumor assessments, quality of life assessment)
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation
- AE start and end dates
- Date last known to be alive collected on the eCRF form "Subject Status / Survival Follow-up"
- Study drug start and end dates
- Randomization date
- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up).

Assessment dates after the cut-off date will not be applied to derive the last alive date.

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

- If the date is missing it will be imputed as day after date of last contact from the eCRF "Subject Status / Survival Follow-Up" page
- If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) day after the date of last contact and the following:
    - Missing day: 1st day of the month and year of death
    - Missing day and month: January 1st of the year of death

For **tumor assessments**, if there are multiple scan dates associated with an evaluation, i.e., radiological assessments occur over a series of days rather than the same day, the choice of date of assessment could impact the date of progression and/or date of response. If there are multiple scan dates associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.

Handling of multiple HRQoL and/or Laboratory records per nominal visit

- The record to be included in summary statistics will be the one which is the nearest to the scheduled visit.
- In case there are records with equal distance between assessment date and scheduled visit (i.e. one before, one after), the one after nominal visit will be included in summary statistics for HRQoL and the one prior to the nominal visit day will be used for laboratory assessments.
- Over the course of the study, there have been cases observed where two or more results for a given analyte share visit name, collection date, and collection time. In such cases, an additional record for the given parameter, visit, date, and time is output with DTYPE = 'AVERAGE' and AVAL = the average of the results from the duplicate records. When a duplicate record could be chosen for the Baseline record, or for inclusion in by-visit summaries, the DTYPE = 'AVERAGE' record is used.
- All available assessments will be kept in ADaM dataset (ADQS and ADLB(HEMA/CHEM)) as all individual assessments may be included in some of the analyses (e.g. TUDD analysis).

All statistical analyses will be performed using SAS® Version 9.4 or higher, or R (www.r-project.org), version 3.2.5 or higher.

Laboratory values outside of limits of quantitation

For laboratory assessments indicating that the sample tested below the lower limit of quantitation (LLOQ) for the analyte in question, half the lower limit of quantitation (½*LLOQ) will be used as the analysis value. For laboratory assessments testing above the upper limit of quantitation (ULOQ), the analysis value will be imputed as ULOQ + 1*(the level of precision of ULOQ). For example, if a value ">7.0" is reported, the level of precision is 0.1 and the analysis value will be 7.1. The imputed analysis values will be used in all analyses (e.g. summary statistics) and will be used for determination of toxicity grading. Listings will present the standardized unit values as provided in the raw data, indicating the assessment value was not quantifiable.

11 Trial Subjects

The subsections in this section include specifications for reporting subject disposition and treatment/trial discontinuations. Additionally procedures for reporting protocol deviations are provided.

11.1 Protocol Deviations

Analysis sets: FAS / mFAS by PD-L1 expression status / Safety Cohort Analysis Set

All important protocol deviations (IPDs) according to ICH E3 will be reported. These include:

- Subjects that are dosed on the study despite not satisfying the inclusion and exclusion criteria;
- Subjects that develop withdrawal criteria whilst on the study but are not withdrawn;
- Subjects that receive the wrong treatment or an incorrect dose;
- Subjects that receive an excluded concomitant medication;
- Deviation from Good Clinical Practice (GCP).

IPDs will be determined for all subjects by either medical review processes or programming based on the inclusion/exclusion criteria or other criteria presented in the protocol.

In addition, a subset of IPDs are identified as clinically important protocol deviations (CIPD) and are included in Appendix II.

All important protocol deviations will be documented in CDISC datasets whether identified through site monitoring, medical review or programming. The complete list of IPDs and CIPDs are maintained by the medical team and will be finalized prior to database lock. These protocol deviations will be presented in a data listing.

All IPDs and CIPDs identified by medical review process and/or programming will be presented in the summary tables by treatment arm and in data listings.

12 Demographics and Other Baseline Characteristics

Analysis sets: FAS / mFAS, all by PD-L1 expression status, Safety-AS (including Safety Cohort column), modified Safety-AS

12.1 Demographics

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

- Demographic characteristics
  - Gender: Male, Female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown

- Ethnic origin: Hispanic/Latino (Yes/No), Japanese (Yes/No)
- Age (years): summary statistics
- Age categories : <65 years, ≥65 years
  - 65-<75, 75-<85, ≥85
- Pooled Region:
  - North America and Western Europe
  - Eastern Europe
  - Asia
  - Rest of the World
- Geographic Region:
  - North America
  - Latin America
  - Western Europe
  - Eastern Europe
  - Australasia
  - Asia
  - Africa
  - Middle East
- EEA: Yes or No
- Eastern Cooperative Oncology Group (ECOG) Performance Status: 0 or 1
- Height (cm)
- Weight (kg)
- Body Mass Index (BMI) (kg/m^2)
- Body Surface Area (BSA) (m^2)

Site codes will be used for the determination of the subject's geographic region.

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

12.2 Medical History

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

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

12.3 Other Baseline Characteristics

12.3.1 Disease Characteristics

Information on disease characteristics collected on the "Disease History" eCRF page will be summarized in total and by treatment arm. Summary statistics will be presented for:

- Site of primary tumor (Upper lobe, left lung / Upper lobe, right lung / Middle lobe, right lung / Lower lobe, left lung / Lower lobe, right lung / Overlapping lesion of lung / Lung, NOS)
- Time since initial cancer diagnosis (months)
- Time since first diagnosis of documented, locally advanced, inoperable or metastatic disease (months)
- Tumor histopathologic / cytologic type
  - Squamous cell
  - Non-squamous cell
    - Adenocarcinoma
    - Bronchoalveolar
    - Large cell
    - Other
- TNM classification at initial diagnosis
- Stage at initial diagnosis per the 7th IASLC and AJCC classifications
- TNM classification at study entry
- Stage at study entry per the 7th IASLC and AJCC classifications
- Smoking history
  - Never smoker vs ever smoker (including further breakdown: regular user / occasional user / former user) as collected in the eCRF
  - Smoking exposure (pack-years): 0, <20, 20-<40, ≥40 and summary statistics
  - Years since quitting: never smoker, current smoker, <5, 5-<10, ≥10 and summary statistics
- Liver metastasis at baseline
  - Present (at least one target or non-target lesion with site = "liver")
  - Absent (no target or non-target lesions with site = "liver")
- Bone metastasis at baseline
  - Present (at least one target or non-target lesion with site = "Bone lesion")
  - Absent (no target or non-target lesions with site = "Bone lesion")

- Visceral metastasis at baseline (see Section 8.2 for definition)
  - Present
  - Absent

Baseline characteristics with respect to physical examinations, ECG and hematology/biochemistry will be part of Section 16 (Safety Evaluation).

Listing of disease history will be provided with all relevant data (as collected on the eCRF page for Disease History) and derived variables used in the above table.

12.3.2 Prior Anti-Cancer Therapies

The number and percentage of subjects in each of the following anti-cancer therapy categories will be tabulated:

- Subjects with at least one type of prior anti-cancer treatment
- Subjects with at least one prior anti-cancer drug therapy
- Subjects with at least one prior anti-cancer radiotherapy
- Subjects with at least one prior anti-cancer surgery

Prior anti-cancer drug therapy will be summarized as follows:

- Number of subjects with at least one prior anti-cancer drug therapy
- Number of subjects with at least one prior anti-cancer drug therapy for locally advanced disease
- Type of prior anti-cancer therapy: Cytotoxic therapy / Endocrine therapy / Monoclonal antibodies therapy / Small molecules / Immunotherapy / Other
- Intent of Therapy: Neo-Adjuvant / Adjuvant / Locally advanced
- Best response: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Not assessable / Unknown / Not applicable. Best response is derived from the last treatment regimen
- The prior anti-cancer drugs will also be extensively detailed with the number and percentage of subjects by the drug class and preferred term in a table. A subject will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. The summary will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used.

The listings of prior anti-cancer treatments and procedures will also be provided as follows. These will include the subject identification number, and all the relevant collected data-fields on the corresponding eCRF pages.

- Listing of prior anti-cancer drug therapies

- Listing of prior anti-cancer radiotherapy
- Listing of prior anti-cancer surgeries

12.3.3 PD-L1 Expression Status and Baseline [redacted]

PD-L1 expression level will be summarized for cut-off values of 1%, 50%, and 80% per the pathology report form (PRF). Tumor associated immune cell and tumor interface status ("Positive", "Negative", "Not present") will also be summarized per the Pathology Report Form (PRF).
The percent of tumor cells at each PD-L1 staining intensity (0, 1+, 2+, and 3+) is also collected on the PRF. The results based on the total percentage of tumor cells at each PD-L1 staining intensity will be summarized based on the categories below:

- Negative: $\geq$ 0% to 1%
- Low: $\geq$ 1% to < 50% at any PD-L1 intensity
- Moderate: $\geq$ 50% to < 80% at any PD-L1 intensity
- High $\geq$ 80% at any PD-L1 intensity
- Non-evaluable

The dates of sample collection for PD-L1 expression analysis will be summarized using the following variables:

- Time from sampling date to date of randomization (months), defined as (randomization date – the date of tissue sampling) / 30.4375
- Number of subjects for whom the sample was collected prior to any prior anti-cancer therapy

13 Previous or Concomitant Medications/Procedures

Analysis sets: FAS / mFAS, all by PD-L1 expression status

13.1 Prior and Concomitant Medications/Procedures

Concomitant medications are medications, other than study medications, which started prior to first dose date of study treatment and continued on on-treatment period as well as those started during on-treatment period.

Prior medications are medications, other than study medications and pre-medications for study drug, which are started before first administration of study treatment.

Prior and concomitant medications will be summarized from the "Concomitant Medications Details" eCRF page. Pre-medications for study drug will also be summarized separately.

In cases where the date values do not allow unequivocal allocation of a medication to concomitant (as opposed to prior) medication, the medication will be considered as concomitant medication.

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

A listing of concomitant medications will be created with the relevant information collected on the "Concomitant Medications Details" eCRF page.

All concurrent procedures, which were undertaken any time during the on-treatment period, will be summarized according to the eCRF page "Concomitant Procedures Details". Number and percentage of subjects with concurrent procedures will be tabulated overall and by reason for procedure as collected in the eCRF page "Concomitant Procedures Details".

A listing of concurrent procedures will be created with the relevant information collected on the "Concomitant Procedures Details" eCRF page.

13.2 Subsequent Anti-Cancer Therapies/Procedures

Anti-cancer treatment after discontinuation will be provided in a data listing with data retrieved from eCRF pages on anti-cancer drug treatments, radiotherapies and surgeries after discontinuation. The earliest date of start of new anti-cancer *drug* therapy will be used for the definition of the on-treatment period and the earliest date of start of new anti-cancer therapy will be used for censoring for efficacy analyses.

Number and percentage of subjects with any anti-cancer treatment after discontinuation or PD (including subjects from chemotherapy arm who received avelumab after confirmation of PD) will be tabulated overall and by type of therapy based on the data collected from the anti-cancer treatment after discontinuation or PD eCRF pages: Avelumab / Cytotoxic therapy / Endocrine therapy / Monoclonal antibodies therapy / Small molecules / Immunotherapy / Other. In addition, the number and percentage of subjects who received subsequent immune therapy as checkpoint inhibitors (CPI) (avelumab, nivolumab, pembrolizumab, lambrolizumab, atezolizumab, durvalumab, tremelimumab or ipilimumab) will be tabulated by treatment arms. The final list of subsequent immune therapies will be provided upon medical review of all subsequent anti-cancer therapies.

Summary statistics will be created for best response across all post study treatments based on the data collected from the "Anti-Cancer Treatment after Discontinuation Details" eCRF page. For subjects who received more than one anti-cancer drug therapy after treatment discontinuation, the best overall response among all anti-cancer drug therapies will be summarized.

Summary of subsequent anti-cancer treatment will include the number and percentage of subjects by ATC Classification level 2 and PT. The same approach as prior and concomitant medications will be applied in presenting the summary table of subsequent anti-cancer treatment.

14 Treatment Compliance and Exposure

Analysis sets: Safety Analysis Set / Safety Cohort Analysis Set

All dosing calculations and summaries will be based on avelumab and chemotherapy administration eCRFs pages.

Subjects randomized to arm A will receive an IV infusion of avelumab at a dose of 10 mg/kg over the duration of 1 hour (-10/+20 minutes) once every 2 weeks (one cycle).

Subjects randomized to arm B (chemotherapy platinum-based doublet) arm will receive IV infusion according to the label instructions and per local administration guidelines over the duration of 1 hour once every 3 weeks (one cycle), for a maximum of 6 cycles. The choice of the chemotherapy administered depends on histology:

- Subjects with NSCLC of squamous histology will receive either:
  - gemcitabine in combination with cisplatin (up to 6 cycles); or
  - gemcitabine in combination with carboplatin (up to 6 cycles); or
  - paclitaxel in combination with carboplatin (up to 6 cycles).
- Subjects with NSCLC of non-squamous histology will receive either:
  - pemetrexed in combination with cisplatin (up to 6 cycles); or
  - pemetrexed in combination with carboplatin (up to 6 cycles); or
  - pemetrexed in combination with cisplatin (up to 4 cycles), followed by pemetrexed maintenance; or
  - pemetrexed in combination with carboplatin (up to 4 cycles), followed by pemetrexed maintenance.

Subjects randomized to arm C will receive an IV infusion of avelumab at a dose of 10 mg/kg over the duration of 1 hour (-10/+20 minutes) every week for 12 consecutive weeks, followed by avelumab at a dose of 10 mg/kg once every 2 weeks.

Analysis of exposure will be based on the calculated actual dose levels (total dose / weight for avelumab, total dose / derived BSA for chemotherapy).

Arm A (avelumab):

For subjects randomized to avelumab in arm A, the dose level is calculated as actual dose administered/weight (mg/kg). The last available weight of the subject on or prior to the day of dosing will be used.

The duration of avelumab treatment (in weeks) during the trial is defined as:

$$\text{duration} = \left(\frac{\text{date of last dose of avelumab} - \text{date of first dose of avelumab} + 14}{7}\right)$$

The cumulative dose (mg/kg) of avelumab per subject in a time period is the sum of the actual dose levels that the subject received within that period (i.e. total dose administered (mg) / weight (kg)).

Each cycle is defined by a 2-week period. The dose intensity and the relative dose intensity of avelumab will be calculated for each subject across all cycles.

The dose intensity (mg/kg/cycle) of avelumab per cycle is defined as:

$$\text{dose intensity} = \left(\frac{\text{cumulative dose of avelumab (mg/kg)}}{(\text{duration of avelumab (in weeks)})/2}\right)$$

The relative dose intensity of avelumab is defined as the actual dose intensity divided by the planned dose as specified in the protocol per cycle and expressed in %:

$$\text{relative dose intensity} = 100 \text{ x} \left(\frac{\text{dose intensity (mg/kg/cycle)}}{10 \text{ (mg/kg)}}\right)$$

Arm B (chemotherapy):

For chemotherapy arm, the number and percentage of subjects per platinum-based doublet will be tabulated as well as the number and percentage of subjects with pemetrexed maintenance.

The duration of chemotherapy treatment (in weeks) during the trial is defined as:

$$\text{duration} = \left(\frac{\textit{date of} \text{ last } \textit{dose of} \text{ chemotherapy} - \text{date of first dose of chemotherapy}}{7}\right)$$

where last chemotherapy = maximum of (date of last dose of chemotherapy product1+ Compdays, date of last dose of chemotherapy product2 + CompDays) where CompDays = 21 for Carboplatin, Cisplatin, Pemetrexed and Paclitaxel; 14 for Gemcitabine.

The subjects who received pemetrexed will be considered as having received a maintenance therapy if the platinum-based chemotherapy (cisplatin or carboplatin) has been definitely stopped and if at least one dose of pemetrexed has been received after the date of last platinum-based chemotherapy dose.

The duration of maintenance therapy (in weeks) for subjects with pemetrexed maintenance is defined as:

$$\text{duration} = \left(\frac{\text{date of last dose pemetrexed maintenance} + 21 - \text{date of first dose pemetrexed maintenance}}{7}\right)$$

where the date of first dose of maintenance therapy is the date of first dose of pemetrexed after the date of last dose of associated platinum-based chemotherapy (cisplatin or carboplatin).

Each cycle is defined by a 3-week period. For analysis purpose, a cycle will be assigned for each administration of at least one chemotherapy product. A complete cycle will be assigned if the platinum-based doublet has been administered as planned.

Arm C (weekly avelumab):

For subjects randomized to avelumab in arm C, the dose level is calculated as actual dose administered/weight (mg/kg). The last available weight of the subject on or prior to the day of dosing will be used.

The duration of avelumab treatment (in weeks) during the trial is defined as:

$$\text{duration (period 1)} = \left(\frac{\text{date of last dose in period 1} - \text{date of first dose of avelumab} + 7}{7}\right)$$

where the last dose in period 1 is the last dose of avelumab before (or equal to) week 12

$$\text{duration (period 2)} = \left(\frac{\text{date of last dose of avelumab} - \text{date of first dose in period 2} + 14}{7}\right)$$

where the first dose in period 2 is the start date of the first 2-week administration interval after week 12

$$\text{total duration} = \left(\frac{\text{date of last dose of avelumab} - \text{date of first dose of avelumab} + A}{7}\right)$$

where A is 7 if treatment is discontinued before week 13, or 14 if treatment is discontinued after week 13

The cumulative dose (mg/kg) of avelumab per subject in a time period is the sum of the actual dose levels that the subject received within that period (i.e. total dose administered (mg) / weight (kg)).

Each cycle is defined by a 2-week period. The dose intensity and the relative dose intensity of avelumab will be calculated for each subject across all cycles.

The dose intensity (mg/kg/cycle) of avelumab per cycle is defined as:

$$\text{dose intensity (period 1)} = \left(\frac{\text{cumulative dose of avelumab in period 1 (mg/kg)}}{\text{duration of avelumab in period 1 (in weeks)}/2}\right)$$

$$\text{dose intensity (period 2)} = \left(\frac{\text{cumulative dose of avelumab in period 2 (mg/kg)}}{\text{duration of avelumab in period 2 (in weeks)}/2}\right)$$

$$\text{total dose intensity} = \left(\frac{\text{total cumulative dose of avelumab (mg/kg)}}{(\text{duration in period 1} + \text{duration in period 2 (in weeks)})/2}\right)$$

The relative dose intensity of avelumab is defined as the actual dose intensity divided by the planned dose as specified in the protocol per cycle and expressed in %:

$$\text{relative dose intensity (period 1)} = 100 \times \left(\frac{\text{dose intensity of period 1 (mg/kg/cycle)}}{20\ \text{(mg/kg)}}\right)$$

$$\text{relative dose intensity (period 2)} = 100 \times \left(\frac{\text{dose intensity of period 2 (mg/kg/cycle)}}{10\ \text{(mg/kg)}}\right)$$

$$\text{total relative dose intensity} = 100 \times \left(\frac{\text{total dose intensity (mg/kg/cycle)}}{\text{10 (mg/kg)}}\right)$$

The following summary tables will be provided:

For avelumab arms:

- Duration of therapy (weeks)
- Total number of infusions received
- Cumulative dose of therapy (mg/kg)
- Dose intensity (mg/kg/cycle)
- Relative dose intensity of therapy (%)

For chemotherapy arm:

- Frequency count of different platinum-based doublet used
- Duration of therapy (weeks)
- Number of cycles and complete cycles
- Number and percentage of subjects with pemetrexed maintenance
- Number of maintenance cycles for subjects with pemetrexed maintenance
- Duration of maintenance therapy for subjects with pemetrexed maintenance (weeks)

Dose Reduction for chemotherapy arm

Dose reductions are identified on the administration eCRF page by the investigator. Number and percentage of subjects with at least one dose reduction as recorded in eCRF as well as a breakdown of dose reductions (1 / 2 / 3 / ≥4) will be summarized for the chemotherapy arm.

Dose Delay

Delays will be derived based on infusion date for the three treatment groups and will be grouped into the following categories based on the deviation of the actual to the planned treatment administration day (relative to the previous treatment administration date):

- No delay (including 1-2 days delays)
- 3-6 days delay
- 7 or more days delay

For example, if one subject receives avelumab on day 1, then the next avelumab administration date will be on day 15; however, if the subject receives avelumab at day 16 or 17, this is considered as 'no delay'.

Number and percentage of subjects with at least one delayed infusion as well as a breakdown of dose delays (1 / 2 / ≥3) and maximum length of delay, i.e. the worst case of delay if subjects have multiple dose delays will be summarized.

For arm C (weekly avelumab), results will be presented on the overall treatment period and also separately on the first 12 weeks of treatment (i.e. weekly administration period of avelumab).

Avelumab infusion rate reductions

Infusion rate reductions as recorded on the eCRF will be used for analysis. Number and percentage of subjects with at least one infusion rate reduction as well as a breakdown of infusion rate reductions (1 / 2 / ≥3) will be summarized for the avelumab treatment groups.

A listing of study drug administration will be created with the information collected on the avelumab / chemotherapy administration details eCRF page.

15 Endpoint Evaluation

The primary endpoints of the trial are the progression free survival (PFS) time and overall survival (OS) time. The trial will be considered to have met its objective if either or both of the primary endpoints are positive.

PFS is defined as the time (in months) from randomization to the date of the first documentation of PD as adjudicated by the IRC, or death due to any cause in the absence of documented PD, whichever occurs first. OS time is defined as the time (in months) from randomization to the date of death, regardless of the actual cause of the subject's death.

The secondary efficacy endpoints include best overall response (BOR) according to RECIST 1.1 and as adjudicated by the IRC. Radiological assessments will be used including adjudicated reviews.

A hierarchical testing strategy will be applied to test superiority of either avelumab regimen vs SOC chemotherapy for the primary endpoints PFS and OS. The FAS will be used for the comparisons of Arm A and Arm B and mFAS will be used for the comparison of Arm C and Arm B:

Hierarchy Step 1: PFS and OS in the high expression PD-L1+ population

Hierarchy Step 2: PFS and OS in the moderate and high expression PD-L1+ population (including high expression PD-L1+ population)

Hierarchy Step 3: OS in the any expression PD-L1+ population

A closed testing procedure will be used with weighted Bonferroni tests according to Hommel, Bretz, Maurer 2007 (1) to account for multiplicity. For the closure test, it holds: an intersection hypothesis H_i can only be rejected if all intersection hypotheses implying H_I can also be rejected by their local tests. Further details are provided in Section 15.1.1 CCI [redacted]

15.1 Primary Endpoint Analyses

The primary statistical comparisons of Arm A vs. Arm B will be performed with the FAS. The primary statistical comparison of Arm C vs. Arm B will be performed with the mFAS to ensure a randomized comparison of treatment arms that is not impacted by the time of randomization.

15.1.1 Hierarchical testing

The following procedure is used:

- All hypotheses comparing Arm C vs Arm B get three times the weight of the hypotheses comparing Arm A vs Arm B, ie, weight ¾ and ¼. Corresponding hypotheses for PFS and OS get different weights, ie, 2/3 for PFS and 1/3 for OS.
- Exception: as long as an hypothesis with PD-L1 high expression (H0_1A,H0_2A,H0_1C,H0_2C) is included in an intersection hypothesis the corresponding hypotheses for PD-L1 moderate and high expression (H0_3A,H0_4A,H0_3C,H0_4C) and all hypotheses for any PD-L1+ (H0_6A,H0_6C) get weight 0, as long as both hypotheses (PFS and OS) with PD-L1 moderate expression (H0_3A,H0_4A and H0_3C,H0_4C) is included in an intersection hypothesis the corresponding OS hypotheses for any PD-L1+ (H0_6A and H0_6C) get weight 0.
- For hypotheses with interim analyses the corresponding significance levels at interim and final analysis are defined based on α-spending approaches using local significance levels according to 1 and 2.
- Sum of the weights for each intersection hypothesis is 1.

The statistical hypotheses are described below. The significance levels noted are the minimum significance levels that can to be used for the corresponding hypothesis. Depending on the already rejected hypothesis according to closed testing procedure above the local significance level might increase.

Hierarchy Step 1: Comparisons for high expression PD-L1+ population

PFS for high expression PD-L1+ in the FAS: Arm A versus B comparison at $\alpha = 0.00417$ or above according to closed testing procedure

$H_{0\_1A}: \lambda_{1A}(t) = \theta_{1h}\lambda_{1B}(t), \theta_{1h} \geq 1$, versus $H_{1\_1A}: \lambda_{1B}(t)=\theta_{h}\lambda_{1C}(t), \theta_{1h} < 1$,

OS for high expression PD-L1+ in the FAS: Arm A versus B comparison at $\alpha = 0.00208$ or above according to closed testing procedure

$H_{0\_2A}: \lambda_{2A}(t) = \theta_{2h}\lambda_{2B}(t), \theta_{2h} \geq 1$, versus $H_{1\_2A}: \lambda_{2A}(t)=\theta_{2h}\lambda_{2B}(t), \theta_{2h} < 1$,

PFS for high expression PD-L1+ in the mFAS: Arm C versus B comparison at $\alpha = 0.0125$ or above according to closed testing procedure

$H_{0\_1C}: \lambda_{1C}(t) = \theta_{1h}\lambda_{1B}(t), \theta_{1h} \geq 1$, versus $H_{1\_1C}: \lambda_{1B}(t)=\theta_{1h}\lambda_{1B}(t), \theta_{1h} < 1$,

OS for high expression PD-L1+ in the mFAS: Arm C versus B comparison at $\alpha = 0.00625$ or above according to closed testing procedure

$H_{0\_2C}: \lambda_{2C}(t) = \theta_{2h}\lambda_{2B}(t), \theta_{2h} \geq 1$, versus $H_{1\_2C}: \lambda_{2C}(t) = \theta_{2h}\lambda_{2B}(t), \theta_{2h} < 1$,

If local significant results are obtained in Hierarchy Step 1, inferential testing will proceed to corresponding hypotheses in Hierarchy Step 2.

Hierarchy Step 2: Comparisons for moderate and high expression PD-L1+ population

PFS for moderate and high expression PD-L1+ in the FAS: Arm A versus B comparison at $\alpha = 0.00417$ or above according to closed testing procedure

$H_{0\_3A}: \lambda_{3A}(t) = \theta_{3m}\lambda_{3B}(t), \theta_{3m} \geq 1$, versus $H_{1\_3A}: \lambda_{3A}(t) = \theta_{3m}\lambda_{3B}(t), \theta_{3m} < 1$,

OS for moderate and high expression PD-L1+ in the FAS: Arm A versus B comparison at $\alpha = 0.00208$ or above according to closed testing procedure

$H_{0\_4A}: \lambda_{4A}(t) = \theta_{4m}\lambda_{4B}(t), \theta_{4m} \geq 1$, versus $H_{1\_4A}: \lambda_{2A}(t) = \theta_{4m}\lambda_{4B}(t), \theta_{4m} < 1$,

PFS for moderate and high expression PD-L1+ in the mFAS: Arm C versus B comparison at $\alpha = 0.00125$ or above according to closed testing procedure

$H_{0\_3C}: \lambda_{3C}(t) = \theta_{3m}\lambda_{3B}(t), \theta_{3m} \geq 1$, versus $H_{1\_3C}: \lambda_{3C}(t) = \theta_{3m}\lambda_{3B}(t), \theta_{3m} < 1$,

OS for moderate and high expression PD-L1+ in the mFAS: Arm C vs B comparison at $\alpha = 0.00625$ or above according to closed testing procedure

$H_{0\_4C}: \lambda_{4C}(t) = \theta_{4m}\lambda_{4B}(t), \theta_{4m} \geq 1$, versus $H_{1\_4C}: \lambda_{4C}(t) = \theta_{4m}\lambda_{4B}(t), \theta_{4m} < 1$,

If local significant results are obtained for Hierarchy 2, inferential testing will proceed to corresponding hypotheses in Hierarchy 3.

Hierarchy Step 3: Comparisons for any expression PD-L1+ population:

OS for any expression PD-L1+ in the FAS: Arm A versus B comparison at $\alpha = 0.00208$ or above according to closed testing procedure

$H_{0\_6A}: \lambda_{6A}(t) = \theta_{6l}\lambda_{6B}(t), \theta_{6l} \geq 1$, versus $H_{1\_6A}: \lambda_{6A}(t) = \theta_{6l}\lambda_{6B}(t), \theta_{6l} < 1$,

OS for any expression PD-L1+ in the mFAS: Arm C vs B comparison at $\alpha = 0.00625$ or above according to closed testing procedure

$H_{0\_6C}: \lambda_{6C}(t) = \theta_{6l}\lambda_{6B}(t), \theta_{6l} \geq 1$, versus $H_{1\_6C}: \lambda_{6C}(t) = \theta_{6l}\lambda_{6B}(t), \theta_{6l} < 1$,

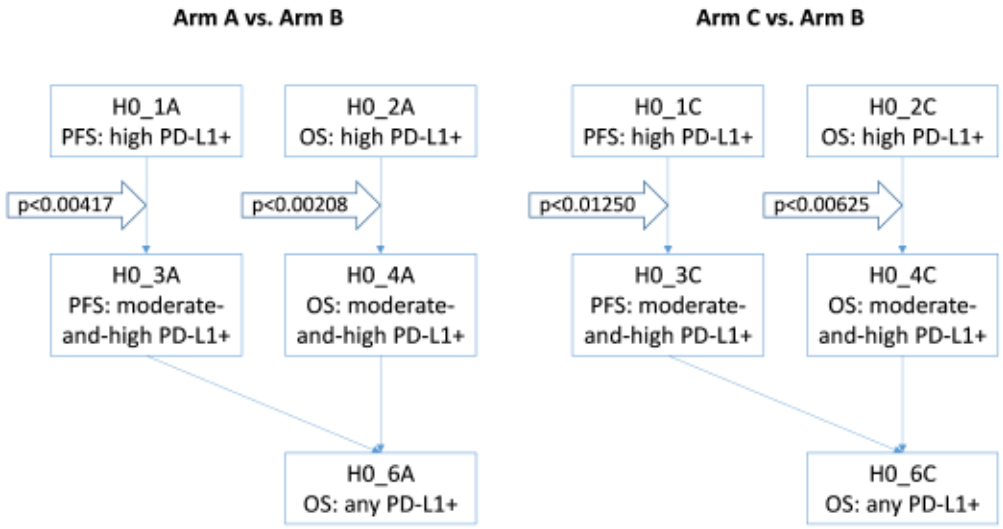


Figure 1 Inferential Testing Procedure

Interim analyses are planned for PFS and OS based on required OS events and minimum follow-up time of 10 months for the high expression PD-L1+ population. Interim and final analysis for PFS and OS will be conducted at the same time. Alpha will be spent according to Lan-DeMets with O'Brien-Fleming-like boundaries with the local significance level at the first hierarchy step. Further details are given in Section 7.2.

15.1.2 Progression Free Survival

Analysis sets: FAS / mFAS, by PD-L1 expression status

PFS is defined as the time from date of randomization until date of the first documentation of PD per RECIST 1.1 or death due to any cause in the absence of documented PD, whichever occurs first. The PFS will be defined according to RECIST 1.1 and as adjudicated by an IRC. Details on determination of the first disease progression date will be specified in the IRC charter.

PFS time (in months) is defined as:

(Date of PD or death - date of randomization + 1) / 30.4375 (months)

For the primary analysis, PFS data will be censored on the date of the last adequate tumor assessment for subjects who do not have an event (PD or death), for subjects who start new anticancer treatment prior to an event, or for subjects with an event after two or more missing tumor assessments. Subjects who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of randomization unless death occurred on or before the time of the second planned tumor assessment, in which case the death will be considered an event.

If a tumor assessment was performed on the same day as start of new anti-cancer therapy, it will be assumed that the tumor assessment was performed prior to the start of the new anti-cancer therapy, therefore the tumor assessment will be included in the derivation of the PFS time.

An adequate tumor assessment is defined as a tumor assessment with result that is not "NE" or "NA".

Some subjects may start new anticancer therapy prior to documented disease progression. For such subjects, the protocol requires tumor assessments to be continued per schedule of assessments through documented disease progression or death to enable conduct of a supportive analysis for PFS that will ignore (that is, not censor subjects at) start of new anticancer treatment.

The censoring and event date options to be considered for the PFS analysis are presented in Table 5.

Table 5 Outcome and event dates for PFS and duration of response analyses

Scenario	Date of event/censoring	Outcome
No baseline assessment	Date of randomization	Censored[a]
PD or death: After at most one missing or inadequate post-baseline tumor assessment, OR ≤ 12 weeks after randomization	Date of PD or death	Event
PD or death after 2 or more missing[c] or inadequate post-baseline tumor assessments	Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
No PD and no death	Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
Treatment discontinuation due to 'Disease progression' without documented progression	Not applicable	Information is ignored. Outcome is derived based on documented progression only.
New anti-cancer therapy given	Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored

[a] However if the patient dies ≤12 weeks after randomization the death is an event with date of death date
[b] If there are no adequate post-baseline assessments prior to the PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria were met the censoring will be on the date of randomization
[c] Derivation of 2 or more missing tumor assessments according to the following rules:

Last Evaluable Scan Prior to Event (PD or Death), days since treatment start	Censor if Duration Between Last Scan and Event
[Day 1, Day 322]	>84 days (12 weeks)
[Day 323, Day 365]	>126 days (18 weeks)
>Day 365	>168 days (24 weeks)

The primary efficacy analysis will compare the PFS time between arm A and arm B (FAS), and between arm C and arm B (mFAS), and will be performed using a one-sided stratified log rank test based on the significance level at the interim and the final analysis as specified in Table 2, Section 7.2 and using the hierarchical approach defined in Section 15.1.1. The stratification factors

are those used for randomization as captured via the IWRS (NSCLC histology: squamous versus non-squamous cell, and PD-L1 tumor expression: low expression versus moderate expression versus high expression). Missing PD-L1 expression classification for subjects enrolled prior to Amendment 3 will be imputed from the results of the Dako PD-L1 immunohistochemistry (IHC) pharmDx companion diagnostic test.

The following null hypotheses will be tested (see Section 15.1.1 for more details):

$$H_0: \lambda_A(t) = \theta\lambda_B(t),\ \theta \geq 1, \text{ versus } H_1: \lambda_A(t)=\theta\lambda_B(t),\ \theta < 1,$$

$$H_0: \lambda_C(t) = \theta\lambda_B(t),\ \theta \geq 1, \text{ versus } H_1: \lambda_C(t)=\theta\lambda_B(t),\ \theta < 1,$$

where $\lambda(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in treatment groups A (avelumab), B (chemotherapy) and C (weekly avelumab).

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio and its two-sided CI in line with the local significance level. In addition 95% CIs will be provided. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), i.e. for the i-th stratum the hazard function is expressed as: $h(i;t) = h(i,0;t)\ \exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment group (0=chemotherapy, 1= avelumab) and β is the unknown regression parameter.

In order to account for the group sequential design as applied in this study, the repeated CI method according to Jennison and Turnbull (2), will be used to construct the two-sided CI for the interim and the final analysis. The alpha value as specified in Table 2, Section 7.2 will be applied to calculate the two-sided CI for the interim analysis and the final analysis, respectively.

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

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median survival time with two-sided 95% CIs. In particular, the PFS rate at 6, 9, 12 and 15 months will be estimated with corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (3) and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (4) (conftype=loglog default option in SAS PROC LIFETEST). The estimate of the standard error will be computed using Greenwood's formula. Kaplan-Meier plots will be displayed by PD-L1 status and by NSCLC histology.

The proportional hazards assumption for the primary analysis will be visually checked by plotting log(-log[survival]) versus log(time) by treatment arm.

For imputing missing parts of dates of death, the rules defined in the Section 10 will be used. In all other cases missing or incomplete dates will not be imputed. Progression dates are expected to be reported as complete dates.

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

- Administrative censoring: Ongoing in the study without an event
- Non-administrative censoring:
  - No baseline assessment
  - No adequate post-baseline assessment
  - Start of new anti-cancer therapy
  - Event after 2 or more missing or inadequate post-baseline assessments
  - Withdrawal of consent
  - Lost to follow-up

Lost to follow-up will include the following subjects:

- Lost to follow-up status is collected on the eCRF treatment termination page prior to the analysis cut-off

Two or more missed or inadequate post-baseline tumor assessments immediately preceding analysis cut-off date (see criteria in Table 5). The PFS time or censoring time and the reasons for censoring will also be presented in a subject listing.

15.1.3 Overall Survival

Analysis sets: FAS / mFAS, by PD-L1 expression status

OS time is defined as the time from randomization to the date of death, regardless of the actual cause of the subject's death. For subjects who are still alive at the time of data analysis or who are lost to follow up, OS time will be censored at the last recorded date that the subject is known to be alive as specified in Section 10 as of the data cut-off date for the analysis.

Overall survival time (in months) is defined as:

(Date of death – date of randomization +1) / 30.4375

The analysis of OS will be analogous to that for PFS time as described in Section 15.1.2. The nominal p-value as specified in Table 2 will be applied to calculate the two-sided CI for the interim analysis and the final analysis, respectively.

Cox's Proportional Hazard model stratified by the randomization strata will be used to calculate the hazard ratio and its two-sided CI. Kaplan-Meier estimates will be presented by treatment group together with a summary of associated statistics including the median survival time with two-sided CIs. In particular, the OS rate at 12, 18, 24 and 30 months will be estimated with corresponding two-sided 95% CI.

Frequency (number and percentage) of subjects 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

Lost to follow-up will include the following subjects:

- Lost to follow-up status is collected on the eCRF treatment termination page prior to the analysis cut-off;
- Subjects with the last alive date > 14 weeks prior to the analysis cut-offdate (duration of 14 weeks is based on the assessment schedule of every 3 months for survival follow-up interval + 1 week window).

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

15.1.4 Sensitivity Analyses of Primary Endpoints

Analysis sets: FAS / mFAS, by PD-L1 expression status

The following sensitivity analyses will be performed to explore the robustness of the primary analysis. These analyses are regarded as purely CCI The unadjusted 95% CI will be calculated for all sensitivity analyses of the primary endpoints. The sensitivity analyses will include the following:

- PFS based on IRC assessment and counting all PD and deaths as events without considering the censoring rule for PD or death as described in Table 5;
- PFS based on IRC assessment and initiation of subsequent anti-cancer therapies will not be used as a censoring reason for PFS;
- PFS and OS based on treated subjects only (safety analysis set)
- An unstratified analysis will be performed to compare the PFS and OS time and to estimate the treatment effect;
- The primary analysis will be repeated using strata according to eCRF instead of IWRS data (for histology only);
- PFS according to investigator assessment will be analyzed. For this sensitivity analysis, the date of tumor assessment will be derived from the "Assessment of Disease Based on Imaging" eCRF page. The earliest tumor assessment date corresponding to the respective visit that is collected on this page will be used for date of disease progression and date of response;
- A summary of concordance of disease progression status between IRC and investigator assessment will be provided including status of "No Event", "Progressive disease" and "Death

as event" as well as the number of cases where PD was assessed at different timepoints by IRC and investigator (shift tables per treatment arm);

- If the actual number of OS events in high expression PD-L1+ patients is 10% more than the planned number of OS events at the interim analysis, i.e. if there are > 143 OS events in high expression PD-L1+ patients for arms B and C based on the prospectively determined cut-off date at the interim analysis or > 191 events at the final analysis, and the increase is not due to the requirement for at least 10 or 20 months of follow-up, the analysis will be repeated using the first cut-off date at which exactly 130 or 173 OS events, respectively, in high expression PD-L1+ patients were observed in the study;
- The primary OS analysis will be repeated by censoring those FAS and mFAS subjects who received subsequent immune therapy such as CPI (avelumab, nivolumab, pembrolizumab, lambrolizumab, atezolizumab, durvalumab, tremelimumab or ipilimumab) after discontinuing the study treatment with the date of the first dose of subsequent anti-cancer therapy minus 1 day. The final list of subsequent immune therapy will be provided upon medical review of all subsequent anti-cancer therapies.
- The interaction term for treatment*PD-L1 status(low/moderate/high) will be assessed for both OS and PFS via Cox regression models. The Cox regression model will be fitted with the PFS or OS time as the dependent variable; PD-L1 expression status, treatment, and the treatment-by-PD-L1 expression status interaction as explanatory variables. A p-value for the interaction test (Likelihood Ratio test) will be provided.

Methods for evaluating the validity of model assumptions

The proportional hazards assumption will be checked visually for the primary analysis by plotting log(-log(PFS or OS, respectively)) versus log(time) within each randomization stratum.

Schoenfeld residuals including a LOESS curve will be plotted to investigate graphically violations of the proportional hazards assumption. Schoenfeld residuals will be computed in SAS using the PHREG procedure and using the OUTPUT statement and the keyword RESSCH. With proportional hazards the LOESS curve should be parallel to the x-axis.

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

Restricted Mean Survival Time (RMST)

The RMST methodology is applicable independently of the proportional hazards (PH) assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis results. In particular, as it pertains to the cut-off point (τ) to evaluate the RMST, it is noted that the cut-off point should not exceed the minimum of the largest observed time for both treatment arms so that the RMST of all treatment arms being evaluated can be adequately estimated and comparison between treatments is feasible; τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval. The RMST up to time τ can then be interpreted as the expected survival time restricted to the common follow-up time τ among all patients. The selection of τ should ensure that the RMST evaluation will not go beyond the maximum time point where the evaluation can be performed

while also taking into account a large period of time that is expected to provide a meaningful assessment of treatment effect. To avoid arbitrary selection of the common cut-off τ for both treatment arms, three sets of analyses will be performed:

- τ_1 = minimum of (largest observed survival time for avelumab arms, largest observed survival time for chemotherapy arm).
- τ_2 = minimum of (largest survival event time for avelumab arms, largest survival event time for chemotherapy arm).
- τ_3 = midpoint between τ_1 and τ_2.

In this section, 'survival' is meant to denote PFS or OS, respectively.

The treatment effect between the avelumab arms and the chemotherapy arm will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and 1-sided p-value will be generated. RMST as a function of τ and the associated treatment effect between the three treatment arms will be plotted against time τ.

CCI [redacted]

Multivariable Cox regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. This analysis will be based on the PFS according to the IRC assessment and OS as defined in the primary analysis (Sections 15.1.2 and 15.1.3) using the censoring rules described in Table 5. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The Cox's Proportional Hazard model is defined as: $h(t) = h(0;t) \exp(Xb)$, where $h(0;t)$ defines the baseline hazard function and X defines the vector of explanatory variables and b the unknown vector of regression parameters. Variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). This analysis will be performed using the stepwise selection method in SAS (Proc PHREG). Once this procedure stops, the factor 'treatment group' will be added to the last selected model in order to evaluate the effect of treatment on PFS or OS time when adjusted for the selected explanatory variables. The hazard ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% confidence intervals. No interactions will be considered. Post-baseline factors will not be considered for the model.

15.1.5 Subgroup Analysis of Primary Endpoints

Analysis sets: FAS / mFAS, by PD-L1 expression status

The subgroup analyses will be performed on the primary endpoints for all subgroup levels defined in Section 8.2 "Subgroup Analysis Sets". All the subgroup analyses are exploratory and will be

performed as unstratified; no adjustment for multiplicity will be performed. The PFS and OS time between the two avelumab groups and the chemotherapy group will be compared using a two-sided unstratified log rank test per subgroup level; and the unstratified HR and its corresponding 95% CI will be computed per subgroup level.

In the case of a low number of subjects within a category (<25 subjects with high PD-L1+ expression from all three treatment groups combined), the categories will be pooled when meaningful.

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

- Model 1: treatment + subgroup
- Model 2: treatment + subgroup + treatment*subgroup-variable

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

The HR and its corresponding 95% CI of all subgroups will also be presented in a forest plot.

This analysis will be done on the FAS and mFAS, respectively.

15.1.6 Time of Follow-Up

Analysis sets: FAS / mFAS, by PD-L1 expression status

A Kaplan-Meier plot for PFS and OS follow-up duration will also be generated to assess the follow-up time in the treatment arms reversing the PFS and OS censoring indicator.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with the median time of follow-up for PFS and OS. In particular, the follow-up rate for OS at 12, 18, 24, and 30 months and the follow-up rate for PFS at 6, 9, 12, and 15 months will be estimated with corresponding two-sided 95% CIs.

15.2 Secondary Endpoint Analyses

Secondary efficacy analyses will be performed on the FAS and the mFAS, by expression PD-L1+ status. Secondary analysis of subject-reported outcomes / quality of life will be performed on the HRQoL analysis set and modified HRQoL analysis set.

15.2.1 Confirmed Best Overall Response

Analysis sets: FAS / mFAS, by PD-L1 expression status

The confirmed BOR is defined as the best response obtained among all tumor assessment visits after the date of randomization until documented disease progression, taking requirements for

confirmation into account as detailed below. The tumor response at each assessment visit will be determined according to RECIST 1.1 (Appendix I) by an IRC. Radiological assessments will be used including adjudicated reviews. Details of determination of tumor response date are provided in Imaging Review charter document.

Only tumor assessments performed before the start of any further anti-cancer treatment will be considered in the assessment of confirmed BOR. Clinical deterioration will not be considered as documented disease progression. If a tumor assessment was performed on the same day as start of new anti-cancer therapy, it will be assumed that the tumor assessment was performed prior to the start of the new anti-cancer therapy, therefore the tumor assessment will be included in the assessment of BOR.

The following requirement is taken into account for determination of confirmed BOR:

- PR or CR needs to be confirmed at a subsequent tumor assessment, preferably at the regularly scheduled 6-week assessment interval, but no sooner than 4 weeks after the initial documentation of CR or PR and before progression;
- PR or CR can be confirmed at an assessment later than the next assessment after the initial documentation of PR or CR, respectively and before progression;
- The minimum duration for a BOR of SD is defined as at least 6 weeks after randomization and before progression;
- PD = progression ≤ 12 weeks after date of randomization (and not qualifying for CR, PR or SD), i.e. tumor assessment of PD that is >12 weeks after date of randomization and there is no tumor assessment in between will have a BOR of NE.

Table 6 summarizes the derivation rules described by Eisenhauer et al. for the confirmed BOR when confirmation from subsequent assessment is needed. For subjects who have non-target lesions only at baseline, the time-point tumor assessment of "Non-CR/non-PD" will be evaluated with the same criteria as SD (minimum criteria for SD duration) in deriving the overall confirmed BOR.

Table 6 Best Overall Response When Confirmation of CR/PR Is Required

Overall response 1st time point	Overall response subsequent time point	Best overall response (BOR)
CR	CR	CR
CR	PR	SD, PD or PR[a]
CR	SD	SD provided minimum criteria for SD duration met; otherwise, PD
CR	PD	SD provided minimum criteria for SD duration met; otherwise, PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise, PD

Overall response 1st time point	Overall response subsequent time point	Best overall response (BOR)
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, NE = non-evaluable.

[a] If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the patient had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

The confirmed objective response rate (ORR) is defined as the proportion of subjects having reached a confirmed BOR of PR or CR according to RECIST 1.1 and as provided by the IRC in the analysis population. Subjects with a BOR of non-CR/non-PD (possible only for subjects without measurable disease at baseline) are not considered as having achieved objective response. Therefore these subjects will only be counted in the denominator of the rate, but not in the numerator.

The confirmed ORR by treatment group will be calculated along with the two-sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option).

The association of treatment arm and objective response will be tested by the General Association Statistic of the Cochran-Mantel-Haenszel test (CMH) (5) with the randomization strata taken into account. The null hypothesis of no association in any of the randomization strata is tested against the alternative, which specifies that there is an association between treatment effect and tumor response at least in one randomization stratum.

The stratified odds ratio in terms of objective response will also be estimated along with its 95% CI to compare the treatment effect. The odds ratio is defined as the odds of objective response with avelumab divided by the odds of objective response with chemotherapy. The Breslow-Day test will be used to check the homogeneity of the odds ratio across the randomization strata. It tests the null hypothesis that odds ratios in all strata are equal against the alternative hypothesis that at least in one stratum the odds ratio is different (6).

In case the null hypothesis of homogeneity of odds ratios across strata is not rejected at the alpha level of 5% two-sided, the common odds ratio will be determined using the Mantel-Haenszel estimate (by the FREQ procedure using CMH option in SAS); if the null hypothesis of homogeneity of odds ratio across all strata is rejected, the odds ratio per stratum will be calculated with the corresponding exact CI (7).

In addition, the number and percentage of subjects with BOR of CR, PR, SD, PD, non-CR/non-PD, and NE will be tabulated. Subjects with BOR of NE will be summarized by reason for having NE status. The following reasons will be used:

- No baseline assessment (independent review committee identifies neither any target nor any non-target lesions)

- No post-baseline assessments due to death within 6 weeks after randomization
- No post-baseline assessments due to other reasons
- All post-baseline assessments have overall response NE
- New anticancer therapy started before first post-baseline assessment
- SD of insufficient duration (<6 weeks after randomization without further evaluable tumor assessment)
- PD too late (i.e. tumor assessment of PD was >12 weeks after date of randomization and there was no tumor assessment in between)

Special and rare cases where BOR is NE due to both early SD and late PD will be classified as 'SD of insufficient duration'.

15.2.2 Health-Related Quality of Life

Analysis sets: HRQoL Analysis Set / modified HRQoL Analysis Set, by PD-L1 expression status

Health-Related Quality of Life (HRQoL) will be assessed by the EuroQOL 5-dimensions questionnaire (EQ-5D) and the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 items (QLQ-C30) supplemented by the Quality of Life Questionnaire-Lung cancer 13-item module (QLQ-LC13). CCI [redacted]

The EQ-5D questionnaire is a generic measure of health status that provides a descriptive profile and a simple index value. It includes 5 items assessing mobility, self-care, usual activities, pain / discomfort, anxiety / depression on a 5-level response scale (EQ-5D-5L): no problems, slight problems, moderate problems, severe problems and extreme problems. It also includes a visual analogue scale ranging from 0 to 100 for self-rated health status. A higher score indicates a better health status.

The EQ-5D-5L scoring system will also be converted into a single index value. The index value is country specific and is a major feature of the EQ-5D instrument, facilitating the calculation of quality-adjusted life years (QALYs) that are used to inform economic evaluations of health care interventions. The UK country specific value set will be used in deriving the index value which ranges from -0.594 (worst health state) to 1.000 (best health state).

The EORTC QLQ-C30 is a questionnaire developed to assess the general aspects of health-related quality of life of cancer patients. It comprises 30 items grouped into 15 multi-item and single-item scales (Table 7). The QLQ-LC13 module comprises 13 lung cancer-specific questions incorporated into one multi-item scale designed to evaluate dyspnea and a series of single items assessing different types of pain, as well as, cough, hemoptysis, dysphagia, sore mouth, alopecia, and peripheral neuropathy.

For both QLQ-C30 and QLQ-LC13 questionnaires, a linear transformation is to be applied for each multi-item scale and single item to standardize the raw score to a range from 0 to 100 as follows:

For functional scales: Final score = (1 – ((RS – 1) / Item range)) x 100

For symptom and Global Health Status/QoL scales: Final score = ((RS – 1) / Item range) x 100

RS is the raw score calculated as the mean of items contributing to the scale. A score of 100 represents the best possible function/quality of life, and the highest burden of symptoms for symptom scales and single items.

Table 7 Scoring rules for EORTC QLQ-C30 and QLQ-LC13

Questionnaire	Type of scale	Scale	Number of items	Items numbers	Item range	Minimum number of non-missing items to calculate the 0-100 score
QLQ-C30	Global Health Status/QoL	Global Health Status/QoL	2	29, 30	6	1
	Functional scales	Physical Functioning	5	1 to 5	3	3
		Role Functioning	2	6, 7	3	1
		Emotional Functioning	4	21 to 24	3	2
		Cognitive Functioning	2	20, 25	3	1
		Social Functioning	2	26, 27	3	1
	Symptom scales / items	Fatigue	3	10, 12, 18	3	2
		Nausea and vomiting	2	14, 15	3	1
		Pain	2	9, 19	3	1
		Dyspnea	1	8	3	1
		Insomnia	1	11	3	1
		Appetite loss	1	13	3	1
		Constipation	1	16	3	1
		Diarrhea	1	17	3	1
		Financial Difficulties	1	28	3	1
QLQ-LC13	Symptom scales / items	Dyspnea	3	3, 4, 5	3	3
		Coughing	1	1	3	1
		Haemoptysis	1	2	3	1
		Sore mouth	1	6	3	1
		Dysphagia	1	7	3	1
		Peripheral neuropathy	1	8	3	1
		Alopecia	1	9	3	1
		Pain in chest	1	10	3	1
		Pain in arm or shoulder	1	11	3	1
		Pain in other parts	1	12	3	1

Questionnaire	Type of scale	Scale	Number of items	Items numbers	Item range	Minimum number of non-missing items to calculate the 0-100 score
		Take any medication for pain	1	13	3	1

For QLQ-C30 module, only Global Health status and functional scales will be included in the analysis. For symptom scale related analysis the Symptom Scales/Items in the QLQ-LC13 module will be used. Unless specifically stated otherwise, all analyses will include data collected on or before the start date for new anti-cancer therapy.

The following analyses will be performed for each of the EQ-5D, QLQ-C30 Global Health status and functional scales and QLQ-LC13 scale:

- Descriptive summary of number of subjects who completed paper HRQoL entry, electronic HRQoL entry and total at each scheduled visit;
- Compliance, and completion rates:

  EQ-5D, QLQ-C30 and QLQ-LC13 assessments of compliance, and completion rates will be summarized for each scheduled administration, considering the following definitions:

$$\%\ Compliance = 100 \times \frac{\text{number of subjects with at least one evaluable HRQoL questionnaire}}{\text{number of subjects in the FAS,MFAS or associated PD—L1 expression subsets}}$$

$$\%Full\ completion\ rate = 100 \times \frac{\text{number of subjects with all items in HRQoL questionniare available}}{\text{number of subjects in the FAS, MFAS or associated PD} - \text{L1 expression subsets}}$$

  The compliance and full completion rate for each questionnaire will be displayed using a line plot with time on the x-axis and completion rate on the y-axis. Separate lines will be presented for each treatment arm.

- HRQoL questionnaires descriptive statistics

  For EQ-5D scores, each of QLQ-C30 Global Health status and functional scores and each of QLQ-LC13 symptoms/items scores, descriptive statistics will be reported for each scheduled visit by treatment group (for visits with $\geq$ 10 subjects in each treatment group). Additionally, change from baseline will also be reported at each scheduled visit by treatment group (for visits with $\geq$ 10 subjects in each treatment group). Descriptive statistics include mean, standard deviation, median, minimum, maximum, and percent missing. Additionally, the percent of subjects at the worst possible score (e.g., 100 in symptom scales and 0 on functional scales and quality of life) and at the best possible score (e.g., 0 in symptom scales and 100 on functional and quality of life) will be reported.

  Change from baseline will also be displayed in a plot with time on the x-axis, mean change score on the y-axis, and standard errors around each mean score. Mean change from baseline will be presented for each treatment arm.

If there are multiple complete assessments for any scheduled visit, the assessment that is closest to the planned visit per protocol will be used in the analysis. If there are multiple complete assessments at Discontinuation visit and End of Treatment visit, the one closest to the treatment discontinuation date will be used in the analysis.

- Longitudinal analysis of change from baseline

  A mixed-effect model repeated measures (MMRM) analysis will evaluate longitudinal change from baseline including measurements after treatment discontinuation on the QLQ-C30 physical function scale and the QLQ-LC13 cough, hemoptysis, dyspnea, and pain in chest scores. These four QLQ-LC13 scores are most prone for changes related to the disease symptom (while other QLQ-LC13 scores are more related to treatment) and therefore are of the most interest in MMRM analysis. Covariates will include the baseline score for the domain score being evaluated and randomization stratification factors, and the questionnaire entry time point including both scheduled and unscheduled visits with $\geq$ 10 subjects in each treatment group will be analyzed as a continuous variable. Two-sided p-values, without adjustment for multiplicity, will be calculated for the difference between the MMRM trajectories for the treatment arms.

- Empirical cumulative distribution function (eCDF)

  For the HRQoL domain scores analyzed using MMRM, an eCDF will display the proportion of subjects who experienced specific changes from baseline to end of treatment. For each eCDF, the change score for each scale will be displayed on the x-axis, and the cumulative percentage of patients achieving that change score or better will be displayed on the y-axis. The eCDF may be interpreted by choosing a change score magnitude that may be considered clinically meaningful on the x-axis and comparing the rate of subjects who achieved at least that amount of change in the treatment arms on the y-axis.

- Time until definitive deterioration (TUDD) analysis

  TUDD analysis will be performed for the QLQ-C30 physical functioning scale and the QLQ-LC13 cough, hemoptysis, dyspnea, and pain in chest scores. The data collected at both scheduled and unscheduled visits will be included in the TUDD analysis.

  TUDD is defined as the time between randomization and the occurrence of definitive deterioration (DD) compared to the baseline score. DD is defined as a HRQoL score change greater than or equal to the Minimal Clinically Important Change (MCIC) compared to baseline and such change is confirmed by no further improvement of HRQoL score as compared to the previous score exceeding MCIC magnitude at next assessment or no further available HRQoL data due to death, occurring within 2 scheduled PRO assessments.

The MCIC is defined in Osaba et.al. (8) as:

- QLQ-C30 physical functioning scale score: A change of 10 normalized score points
- QLQ-C30 or QLQ-LC13 symptom scale: A change of 10 normalized points

The treatment effect will be estimated using a Cox proportional-hazards model stratified by the randomization strata to calculate the hazard ratio and its 95% CI. In addition, the results will be displayed using forest plots, presenting HR and 95% CI for each scale/item.

Kaplan-Meier estimate and its 95% CI will be calculated at the time of scheduled assessment.

Censoring Rules for TUDD Analysis

The censor rules is listed in Table 8.

Table 8: Outcome and event dates for TUDD analyses

Scenario	Date of event/censoring	Outcome
No baseline assessment	Date of randomization	Censored
Baseline but no adequate PRO assessment after baseline	Date of randomization+1	Censored
TUDD event: HRQoL domain score change greater than or equal to the Minimal Clinically Important Change (MCIC) compared to baseline (=TUD event) AND compared to this event, no further improvement of HRQoL domain score exceeding MCIC magnitude at a subsequent assessment, at least 20 days but not more than 84 days later OR no further available HRQoL data due to death, but death occurring within 84 days (=2 scheduled PRO assessments) AND no new anti-cancer therapy administered in-between.	Date of deterioration (first event qualifying as TUDD)	Event
HRQoL domain score change greater than or equal to the Minimal Clinically Important Change (MCIC) compared to baseline AND No subsequent HRQoL assessment available but death did not yet occur or occurred more than 84 days later	Date of last adequate PRO assessment documenting no deterioration before new anti-cancer therapy is given or missed HRQoL assessments	Censor
No TUDD event AND no new anticancer therapy administered	Date of last adequate PRO assessment.	Censor
New anti-cancer therapy administered but no TUDD before start of new anti-cancer therapy	Date of last adequate PRO assessment prior to start of new anti-cancer therapy	Censor

15.2.3 Sensitivity Analyses of Secondary Efficacy Endpoints

Analysis sets: FAS / mFAS, by PD-L1 expression status

For the secondary efficacy endpoint of BOR, the following sensitivity analyses will be performed to assess the robustness of the primary analysis results. These analyses are regarded as purely exploratory. The unadjusted 95% CI will be calculated for all the sensitivity analyses of BOR.

- BOR according to Investigator assessment will be analyzed. For this sensitivity analysis, the date of tumor assessment will be derived from the "Assessment of Disease Based on

Imaging" eCRF page. The earliest tumor assessment date corresponding to the respective visit that is collected on this page will be used for date of disease progression and date of response.

- A multivariable logistic regression analysis will be performed on confirmed BOR based on IERC assessment to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. The subgroup variables defined in Section 8.2 "Subgroup Analysis Sets" will be included in the model. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald test). Once this procedure stops, the factor 'treatment group' will be added to the last selected model in order to evaluate the effect of treatment on BOR when adjusted for the selected explanatory variables. The odd ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% confidence intervals. No interactions will be considered. Post-baseline factors will not be considered for the model.
- A summary of concordance of BOR between IRC and investigator assessment will be provided (shift table by treatment arm).

15.2.4 Subgroup Analysis on Secondary Efficacy Endpoints

Analysis sets: FAS / mFAS, by PD-L1 expression status

Analysis on subgroups as defined in Section 8.2 will be performed for the secondary efficacy endpoint of BOR.

To assess the heterogeneity of treatment effect across the subgroup levels for the secondary endpoint of BOR, a logistic regression model will be fitted with BOR as the dependent variable (=1 for subjects with a confirmed BOR of PR or CR; =0 otherwise); subgroup, treatment, and with and without the treatment-by-subgroup interaction as explanatory variables. A p-value for the interaction term (Wald Chi-Square test) will be provided together with the odds ratio and corresponding 95% CI of the interaction model parameter.

In addition, the ORR along with the two-sided exact Clopper-Pearson 95% CIs will be calculated for each subgroup. The results will be displayed using a forest plot, presenting the odds ratio and 95% CI for each subgroup.

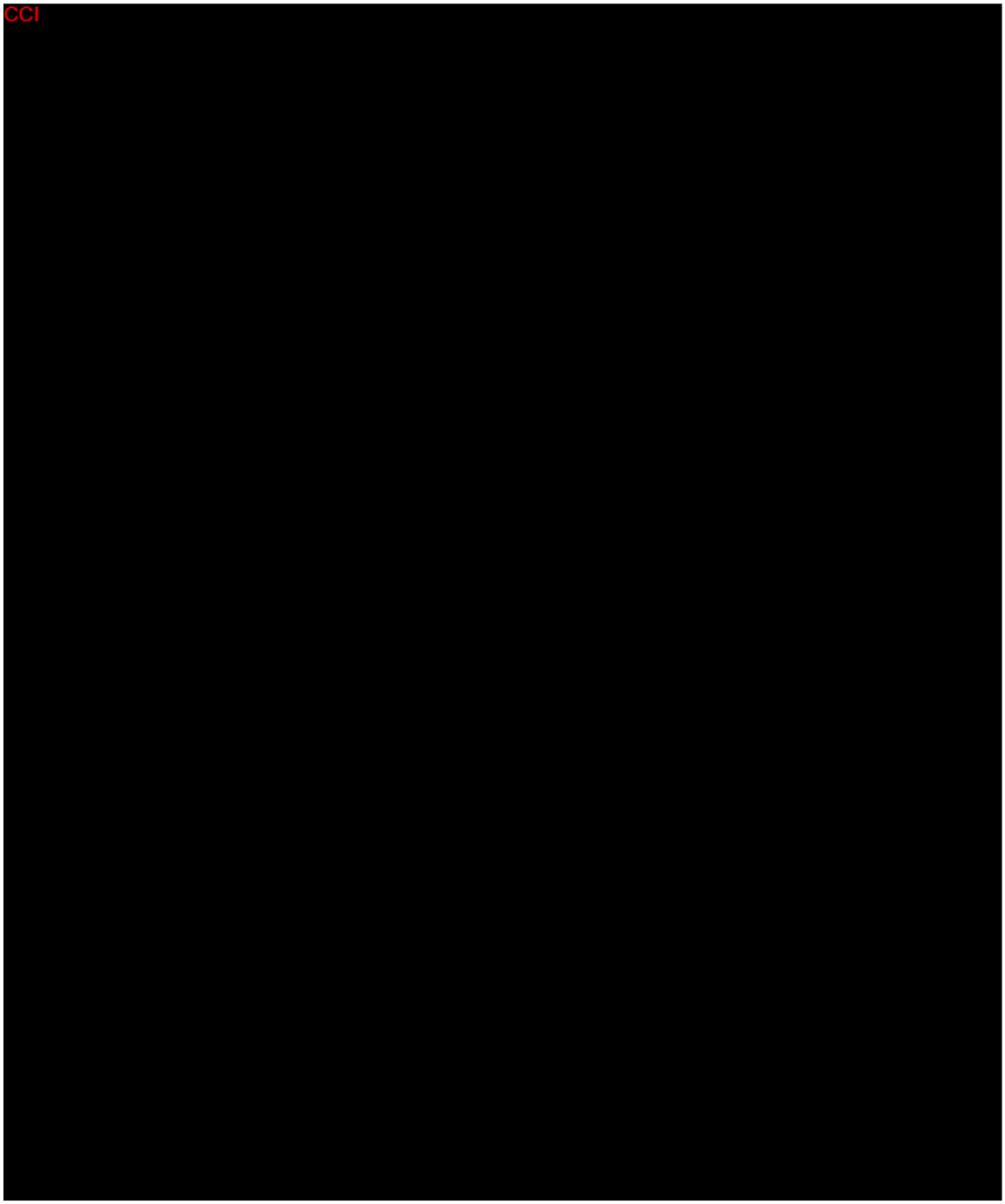
CCI [redacted]

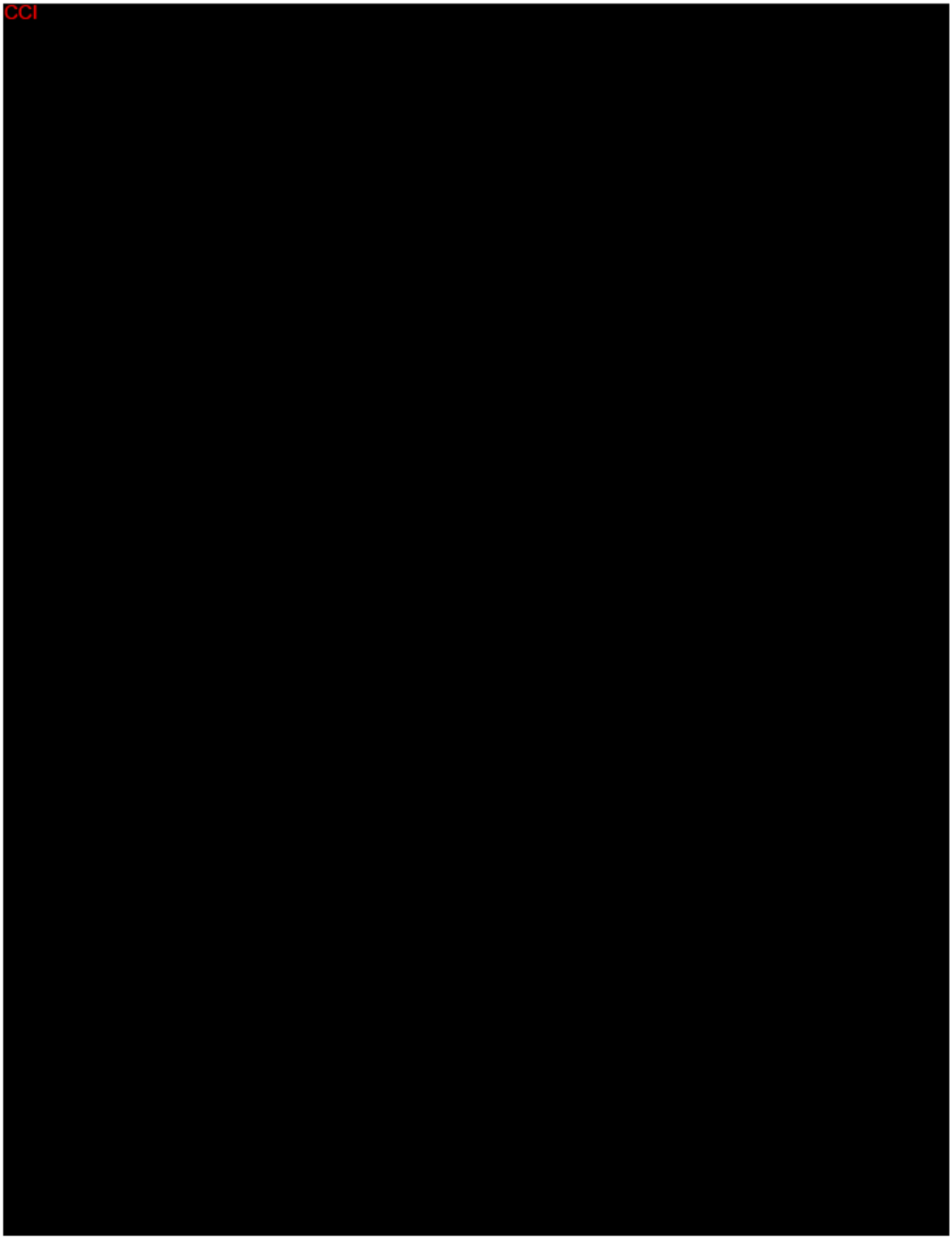
The p-value for the interaction test (Likelihood Ratio test) will be provided together with the HR for OS / odds ratio for ORR and 95% CI of the interaction model parameter. This analysis will be done in the FAS (A versus B) and mFAS (C versus B).

15.3 Other Endpoint Analyses

Analysis sets: FAS / mFAS, by PD-L1 expression status

CCI



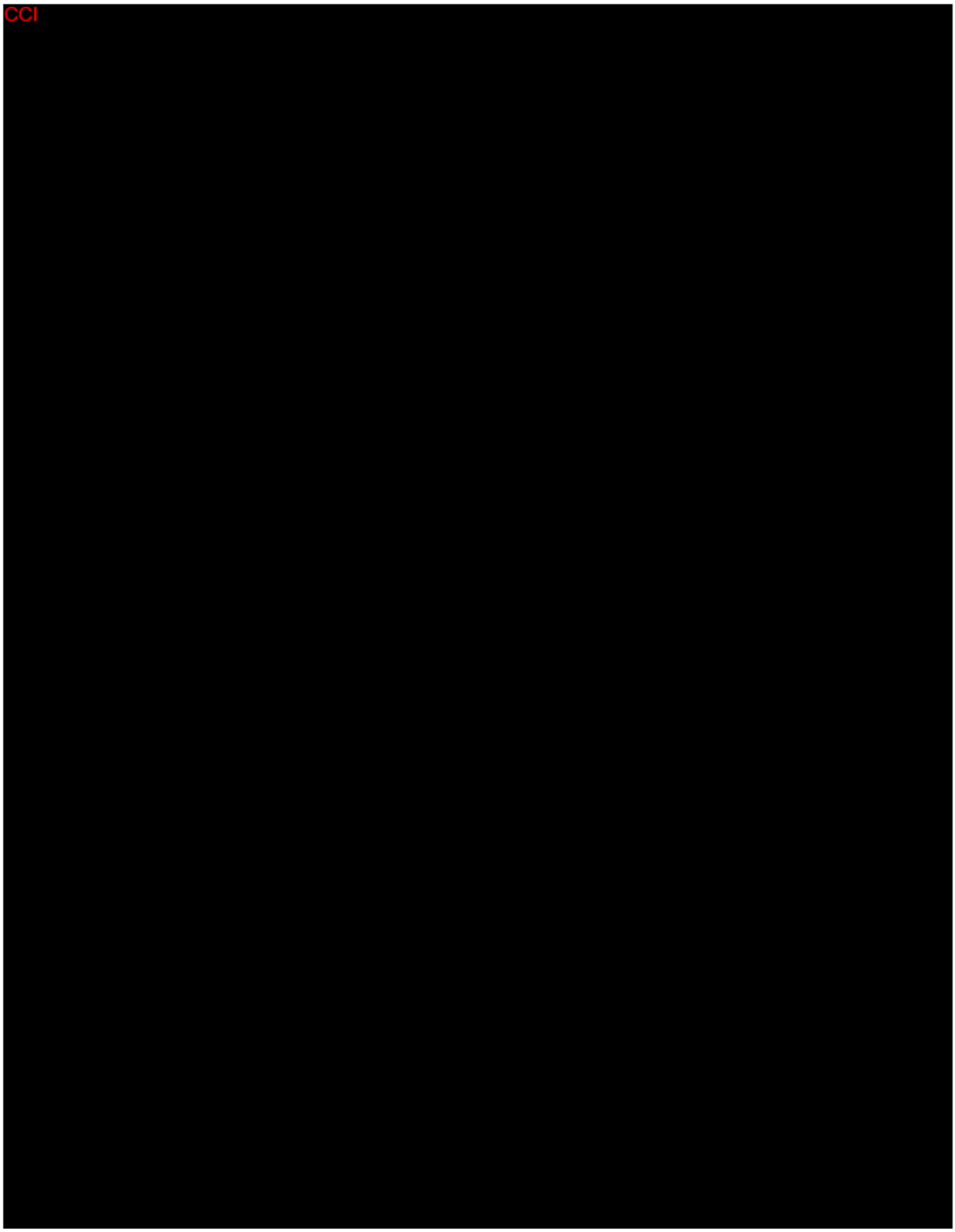

CCI

CCI

[redacted]

CCI

[redacted]


CCI

16 Safety Evaluation

Analysis sets: Safety Analysis Set / modified Safety Analysis Set/ Safety Cohort Analysis Set

The subsections in this section include specifications for summarizing safety endpoints that are common across clinical trials such as adverse events, laboratory tests and vital signs.

Safety analyses will be performed on the Safety analysis set or modified Safety analysis set according to the as-treated principle. The safety endpoints will be tabulated using descriptive statistics.

16.1 Adverse Events

The 3-tier approach is a systematic way to summarize and analyze adverse events (AEs) in clinical studies. AEs in different tiers are analyzed using different levels of statistical analyses.

Analysis (Analysis Population)	Safety Tier	Derivation	Statistical Analysis Methods
Secondary endpoint: Occurrence of TEAEs, treatment-related AEs and serious AEs			
Secondary (SAF)	Tier 1	The AEs identified for Tier 1 reporting in this study are the composite terms of pneumonia and lower respiratory tract infections as defined in Table 10. Composites are based on selected PTs from the MedDRA HLTs "Bacterial lower respiratory tract infections", "Fungal lower respiratory tract infections", "Lower respiratory tract infections NEC", "Respiratory tract infections NEC", "Viral lower respiratory infections" and "Lower respiratory tract and lung infections".	The Tier 1 and Tier 2 AEs will be assessed with a 95% CI for between-group comparisons. For the difference in incidence proportions, the CIs will be based on the method proposed by Miettinen & Nurminen (9). While analyses will be done for all Tier 2 AEs, only AEs with incidence proportion [≥5%] in at least one group will be presented in the body of clinical study report. No multiplicity adjustment will be applied. For this trial, difference in EAIR will also be summarized as exposure periods are expected to show large difference between study intervention groups or between participants within a study intervention group. The CIs will be based on the Wald-type method proposed by Liu et al. (10).
	Tier 2	AEs will be further classified into Tier 2 or Tier 3 (including AEs in Tier 1 being too rare) based on the Rule-of-4. If there are 4 or more participants with the reported term in any study intervention group, that term will be included in Tier 2. Otherwise, it will be included in Tier 3.	
	Tier 3		The Tier 3 AEs will be assessed via summary statistics and risk differences. For this trial, difference in EAIR will also be summarized as exposure periods are expected to show large difference between study intervention groups.
EAIR: Exposure Adjusted Incidence Rate (definition provided below in section)			

Table 10 Categories and Preferred Terms for Tier 1 Events

Category	MedDRA Terms
Pneumonia	Atypical pneumonia, Pneumonia, Pneumonia cytomegaloviral, Pneumonia pneumococcal, Pneumonia respiratory syncytial viral, Pneumonia influenzal, Pneumonia parainfluenzae viral, Pneumonia staphylococcal, Pneumonia streptococcal, Pneumonia viral, Pneumonia necrotising, Pneumonia bacterial, Lung abscess, Pulmonary sepsis, Bronchopulmonary aspergillosis, Pneumonia fungal,Haemorrhagic pneumonia, and COVID-19 pneumonia
Lower respiratory tract infections	Bronchitis, Lower respiratory tract infection, Bronchitis viral, Respiratory tract infection fungal, Respiratory tract infection bacterial, Bronchitis bacterial, Bronchitis fungal, Bronchitis pneumococcal, Severe acute respiratory syndrome, Respiratory tract infection viral, Respiratory tract infection, Lower respiratory tract infection bacterial, Lower respiratory tract infection fungal and Lower respiratory tract infection viral

MedDRA version 24.0

Treatment emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period or if the worsening of an event is during the on-treatment period as defined in Section 10.

All analyses described in Sections 16.1 and 16.2 will be based on TEAEs if not otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not). AEs outside the on-treatment period will be flagged in the listings.

A separate listing will also be provided for AEs that started during second period of avelumab treatment (for subjects in the avelumab arm who re-initiated treatment after PD)

The following categories of TEAEs are defined for reporting:

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (i.e. no answer to the question "Relationship with study treatment").
- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious Adverse Event = Yes).
- **Adverse Events Leading to Treatment Discontinuation:** adverse events leading to permanent discontinuation of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug withdrawn).
- **Adverse Events Leading to Death:** adverse event leading to death (as recorded on the AE eCRF page, Outcome = Fatal, as well as AEs of Grade 5).
- **Immune Related Adverse Events (irAE):** irAEs are identified based on a list of MedDRA PTs and other qualifying criteria (for details see Appendix V).

- **Infusion Related Reaction (IRR):** IRRs are identified based on a list of MedDRA PTs. The detailed criteria of the timing relationship to infusion are specified in Table 16 in Appendix V.

For the chemotherapy arm, an adverse event will be considered as a related adverse event if it is related to at least one of the treatments in the arm. In the same way, an adverse event will be considered as leading to treatment discontinuation if it led to the discontinuation of at least one chemotherapy treatment.

Unless otherwise specified, AEs will be displayed in terms of frequency tables by treatment group by descending frequency of primary system organ class (SOC) and descending frequency of preferred term (PT) in avelumab arm A within SOCs. When the frequency in the SOC or PT is the same in the avelumab Arm A treatment group, the SOC or PT will be displayed in alphabetical order.

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

For analysis purposes, AE related to study treatment are defined as those AE with relationship to study treatment as reported as "related", missing or unknown on the "Adverse Events Details" eCRF page.

Changes in toxicity grade, seriousness or outcome of AEs are recorded as separate entries in the eCRF with associated end and start dates (start date equals end date of previous entry). Such entries reporting the same event in such immediately consecutive periods will be considered as one event in the analysis in case of an improvement in toxicity grade. These events will be kept as separate records in the database in order to maintain the full detailed history of the events. The start date of the initial record in the sequence is taken as start date of the entire event. Duration of the AE and the TEAE flag will be adjusted accordingly in the analysis.

16.1.1 All Adverse Events

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

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

The following frequency tables will be created:

- The overall summary of AEs table will include the frequency (number and percentage) of subjects with each of the following by treatment group:
  - TEAEs
  - TEAEs, grade ≥ 3

  - Related TEAEs
  - Related TEAEs, grade ≥ 3
  - TEAEs leading to permanent treatment discontinuation
  - Related TEAEs leading to permanent treatment discontinuation
  - Serious TEAEs
  - Related serious TEAEs
  - TEAEs leading to death
  - Related TEAEs leading to death
  - Treatment-emergent potential irAEs
  - Treatment-emergent irAEs (for avelumab treatment arm only)
  - Treatment-emergent IRRs
  - Related treatment-emergent IRRs
- TEAEs by SOC and PT and worst grade
- Related TEAEs by SOC and PT and worst grade
- TEAEs leading to death by SOC and PT
- Related TEAEs leading to death by SOC and PT
- TEAEs by SOC and PT: displaying in separate columns the All TEAEs / Related TEAEs / Grade ≥3 TEAEs / Related Grade ≥3 TEAEs
- TEAEs Excluding SAEs with frequency ≥ 5% in any treatment arm by SOC and PT

For subjects who switch to avelumab using the corresponding on-treatment period definition, a separate overall summary of AEs table will be provided for AEs occurring in this period. As long as less than 20 subjects in the study are available in this subset, only a listing will be provided. At the time when at least 20 subjects have switched to avelumab, the summary of AEs table will also be provided, with only a 'switch to avelumab' column.

Adverse events will be illustrated in separate tables by tier according to the 3-tier approach. AEs in all tiers will be sorted in descending order according to the risk difference.

The following tables will be provided in addition also for the mSafety-AS:

- Over all summary of AEs as described above
- TEAEs by SOC and PT: displaying in separate columns the All TEAEs / Related TEAEs / Grade ≥3 TEAEs / Related Grade ≥3 TEAEs
- TEAEs leading to death by SOC and PT
- Tables by tier according to the 3-tier approach (only Arm B vs. Arm C)

Exposure Adjusted Incidence Rate

Exposure Adjusted Incidence Rates (EAIRs) are calculated as number of participants with an AE divided by the sum of the individual times at risk for the first occurrence of an AE of all participants in the safety population from start of study intervention to first onset of AE, end of treatment period, cut-off or death, whichever occurs first.

Exposure will be calculated as time at risk:

- For a subject with AE: study day of first event on study intervention – study day start of study intervention + 1
- Subjects without AE: study day of end of treatment period – study day start of study intervention + 1

EAIRs of first occurrence of an AE of toxicity grade >= 3 is derived in a similar manner.

The following tables will be provided:

- Exposure adjusted incidence rates of AEs by System Organ Class and Preferred Term
- Exposure adjusted incidence rates of related AEs by System Organ Class and Preferred Term
- Exposure adjusted incidence rates of NCI-CTCAE toxicity grade ≥3 AEs by System Organ Class and Preferred Term
- Exposure adjusted incidence rates of related [NCI-CTCAE toxicity grade ≥3] AEs by System Organ Class and Preferred Term
- Exposure adjusted incidence rates of NCI-CTCAE toxicity grade ≥4 AEs by System Organ Class and Preferred Term
- Exposure adjusted incidence rates of related [NCI-CTCAE toxicity grade ≥4] AEs by System Organ Class and Preferred Term

16.1.2 Adverse Events Leading to Permanent Treatment Discontinuation

The following overall frequency tables will be prepared for the adverse event actions that lead to permanent treatment discontinuation (drug withdrawal):

- TEAEs leading to treatment discontinuation by SOC and PT
- Related TEAEs leading to treatment discontinuation by SOC and PT

The following table will be provided in addition also for the mSafety-AS:

- TEAEs leading to treatment discontinuation by SOC and PT

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

16.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

16.2.1 Deaths

All deaths, deaths within 30 days after last dose of study drug, death within 60 days after first dose as well as the primary reason for death, will be tabulated based on information from the "Report of Subject Death" and "Survival Follow-Up" eCRFs.

- Number of Deaths
- Number of Deaths within 30 days after last dose of study treatment
- Number of Deaths within 60 days after first dose of study treatment
- Primary Reason of Death
  - Disease progression
  - Adverse event related to study treatment
  - Adverse event not related to study treatment
  - Other
  - Unknown

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

- AEs with fatal outcome (list preferred terms of AEs with outcome=Fatal, as well as AEs of Grade 5),
- Flag for death within 30 days of last study treatment
- Flag for death within 60 days of first study treatment

16.2.2 Serious Adverse Events

The following overall frequency tables will be prepared for treatment-emergent serious adverse events (SAEs):

- Serious AEs by SOC and PT
- Related serious AEs by SOC and PT

The following table will be provided in addition also for the mSafety-AS:

- Serious AEs by SOC and PT

The listings of SAEs will also be provided with the relevant information.

16.2.3 Other Significant Adverse Event

16.2.3.1 Immune-Related Adverse Events

The following tables will be created for treatment-emergent irAEs by treatment group. Separate summary tables will present potential irAEs (those identified solely by MedDRA preferred term) and algorithmically-assessed irAEs (which meet the criteria outlined in Appendix V). For the potential irAE tables, all three treatment arms will be presented with counts, as applicable. Only the counts for the avelumab arm will be displayed in the classified irAE summary tables. "Cluster" is a compilation of PTs that are categorized by immune-related event of special interest.

- The overall summary of treatment emergent irAEs table will include the following categories:
  - All irAEs
  - Serious irAEs
  - irAEs, grade ≥ 3
  - irAEs leading to permanent treatment discontinuation
  - irAEs leading to death
- irAEs leading to death, by Cluster and PT
- irAEs by Cluster and PT
- irAEs, grade ≥ 3, by Cluster and PT
- irAEs leading to permanent treatment discontinuation by Cluster and PT
- irAEs by Cluster, PT, and worst grade

The listing of all potential irAEs will also be provided with all relevant information including a flag for irAEs with onset outside of the on-treatment period. The following information will be provided for each irAE:

- The time from first study treatment to start of irAE
- The time from most recent study treatment until start of irAE
- The duration of the irAE

The irAEs identified per by the algorithm defined by study medical personnel that are only applicable to avelumab treatment arms will also be flagged in the listing. A separate listing of irAEs with onset after the on-treatment period will also be provided.

16.2.3.2 Infusion Related Reactions

The overall summary of treatment emergent IRR table will include the following summaries:

- Treatment emergent IRR

- Related treatment emergent IRR
- Serious treatment emergent IRR
- Related serious treatment emergent IRR
- Treatment emergent IRR, grade ≥ 3
- Related treatment emergent IRR, grade ≥ 3
- Treatment emergent IRR leading to death
- Related treatment emergent IRR leading to death
- Treatment emergent IRR leading to permanent treatment discontinuation
- Related treatment emergent IRR leading to permanent treatment discontinuation
- Time related to first onset of an IRR (cycle 1/ cycle 2/ cycle 3/ cycle 4 or later). The events should be assigned to the actual drug infusions that the subject received, not to the planned dates. An IRR is assigned to a drug cycle if its onset is at the same date (but not before dosing) or the following day of drug infusion. For avelumab arms, a cycle will be assigned for each avelumab infusion. For chemotherapy arm, a cycle will be assigned for each administration of at least one chemotherapy product.

The following overall frequency tables will be prepared for IRRs:

- IRRs leading to death, by PT
- Related IRRs leading to death, by PT
- IRRs, by PT
- IRRs, Grade ≥ 3, by PT
- Related IRRs, by PT
- Related IRRs, Grade ≥ 3, by PT
- IRRs leading to permanent treatment discontinuation, by PT
- Related IRRs leading to permanent treatment discontinuation, by PT
- Serious IRRs, by PT
- Related serious IRRs, by PT

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

16.2.4 Immunogenicity Subgroup Analysis of Adverse Events

Analysis sets: Safety analysis set, avelumab arms only

A listing of immunogenicity data and TEAEs will be provided containing subject ID, treatment arm, age, gender, study treatment start and stop date, all dates with positive ADA result, all dates

with positive nAb results, preferred term of TEAE, TEAE start date, stop date, CTCAE toxicity grade and flags for immune-related adverse event or infusion related reaction or reason for permanent treatment discontinuation, as well as ADA status group.

Only subjects which are pre-existing positive, transient treatment-emergent positive, or persistent treatment-emergent positive will be listed).

16.3 Clinical Laboratory Evaluation

Analysis sets: Safety analysis set

16.3.1 Hematology and Chemistry Parameters

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

If both central and local labs are collected for a subject, any summary statistics by visit will be based only on the central lab results, while summaries of worst on-treatment grade will be based on both central and local lab results.

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

The worst grade during the on-treatment period will be summarized by treatment group considering only subjects with post baseline laboratory samples: Laboratory tests by NCI-CTCAE grade (0, 1, 2, 3, 4).

Separate laboratory results tables for subjects who switch to avelumab using the corresponding on-treatment period definition will be provided for samples taken during this period (with only a 'switch to avelumab' column), when at least 20 subjects have switched to avelumab. If less than 20 subjects, respective laboratory results will be presented in a listing.

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

For **calcium**, CTCAE grading is based on Corrected Calcium. Corrected Calcium is calculated from Albumin and Calcium as follows based on the International System of Units (SI):

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

Liver function tests:

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

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

- ALT: ≥3×ULN / ≥5×ULN / ≥10×ULN / ≥20×ULN.
- AST: ≥3×ULN / ≥5×ULN / ≥10×ULN / ≥20×ULN.
- (ALT or AST): ≥3×ULN / ≥5×ULN / ≥10×ULN / ≥20×ULN.
- Total bilirubin: ≥2×ULN.
- ALT ≥3×ULN concurrently with total bilirubin ≥2×ULN.
- AST ≥3×ULN concurrently with total bilirubin ≥2×ULN.
- (ALT or AST) ≥3×ULN concurrently with total bilirubin ≥2×ULN.
- (ALT or AST) ≥3×ULN concurrently with total bilirubin ≥2×ULN and ALP >2×ULN.
- (ALT or AST) ≥3×ULN concurrently with total bilirubin ≥2×ULN and (ALP ≤2×ULN or missing).

Concurrent measurements are those occurring on the same date.

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

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

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

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

Parameters with NCI-CTCAE grades available:

The laboratory toxicities will be tabulated using descriptive statistics (count and percentage) during the on-treatment period. The summary statistics will be based on subjects who have at least one post-baseline laboratory assessment.

- The summary of laboratory parameters by CTCAE grade table will include number and percentage of subjects with the worst on-treatment Grade >=0, Grade >=3, and Grade >= 4 during the on-treatment period.
- The shift table will summarize baseline CTCAE grade versus the worst on-treatment CTCAE grade. The highest CTCAE grade during the on-treatment period is considered as the worst grade for the summary.

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

- Hematology:

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

- Serum Chemistry:

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

Parameters with NCI-CTCAE grades not available:

Hematology and chemistry evaluations which can't be graded per CTCAE criteria will be summarized as:

- Number and percentage of subjects with shifts from baseline normal to at least one result above normal during on-treatment period
- Number and percentage of subjects with shifts from baseline normal to at least one result below normal during on-treatment period

In this study, these apply to the following parameters:

- Hematology:

Hematocrit, Erythrocytes, Monocytes, Eosinophils, Basophils, Reticulocytes, Ery. Mean Corpuscular Hemoglobin, Ery. Mean Corpuscular Volume, Ery. Mean Corpuscular Hemoglobin Concentration.

- Serum Chemistry:

  Chloride, C-Reactive Protein, Lactate Dehydrogenase (LDH), Total Protein, Urate.

16.3.2 Other Laboratory Parameters

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

- Coagulation: activated partial thromboplastin time (aPTT) and prothrombin time (INR).
- Urinalysis: all urinalysis parameters
- Other parameters: hormone, and immunology parameters
- Pregnancy test

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

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

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

16.4 Vital Signs

Analysis sets: Safety analysis set

Vital sign summaries will include all vital sign assessments from the on-treatment period. All vital sign assessments will be listed, and those collected outside the on-treatment period will be flagged in the listing, with different flags for those collected during the second avelumab on-treatment period (i.e. switch to avelumab) and those collected outside the second avelumab on-treatment period.

The maximum changes of vital sign measurements from screening/baseline to maximum changes after randomization will be grouped as follows:

Parameters	Categories of Change from Baseline
Body temperature increase	<1°C , 1-<2°C , 2-<3°C, ≥3 °C
Weight increase	<10%, ≥ 10%

Parameters	Categories of Change from Baseline
Weight decrease	<10%, ≥ 10%
Heart rate increase from baseline <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
Heart rate decrease from baseline <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
SBP increase from baseline <140 mmHg; ≥ 140 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
SBP decrease from baseline <140 mmHg; ≥ 140 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP increase from baseline <90 mmHg; ≥ 90 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP decrease from baseline <90 mmHg; ≥ 90 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
Respiration rate increase from baseline <20 bpm ; ≥ 20 bpm	≤5 bpm, >5 – 10 bpm, >10 bpm
Respiration rate decrease from baseline <20 bpm ; ≥ 20 bpm	≤5 bpm, >5 – 10 bpm, >10 bpm

For each patient the worst on-treatment value will be calculated. Missing values will define a separate category.

The following summaries will be prepared for vital sign parameters as grouped above by treatment arm considering only subjects with post baseline values:

- Maximal Shifts (changes in categories)

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

An additional subject data listing will present all changes from baseline reported in the highest categories.

16.5 Other Safety or Tolerability Evaluations

16.5.1 ECG

Analysis sets: Safety analysis set

The 12-lead Electrocardiogram (ECG) assessment will be performed during screening (baseline) and at the Discontinuation / End-of-Treatment visit. For each of the ECG parameters, descriptive statistics at baseline and at the Discontinuation / End-of-Treatment visit and changes from baseline will be presented.

The incidence and percentage of subjects with potentially clinically significant abnormalities (PCSA) for ECG parameters will be summarized during the on-treatment period (including

unscheduled measurements) as defined in Section 10, as well as for subjects who switched to avelumab using the corresponding on-treatment period definition (when there are more than 20 subjects who switched to avelumab). Each subject will be counted only once within each category. As ECG assessments are planned to be performed at screening and Discontinuation/End-of-Treatment visits, the denominator to calculate percentages for each PCSA category is the number of subjects with a Discontinuation/End of Treatment visit. The PCSA criteria are provided in Table 11 below.

Table 11 Potentially Clinically Significant Abnormalities criteria for ECG

Test	Potentially Clinically Significant Abnormalities (PCSA) Criteria
Heart Rate (HR)	≤ 50 bpm and decrease from baseline ≥ 20 bpm ≥ 120 bpm and increase from baseline ≥ 20 bpm
PR Interval	≥ 220 ms and increase from baseline ≥ 20 ms
QRS	≥ 120 ms
QTcF and QTcB absolute	>450 ms >480 ms >500 ms
QTcF and QTcB change from baseline	Increase from baseline > 30 ms Increase from baseline > 60 ms

QT will be corrected based on Fridericia's formula (QTcF= QT/ $\sqrt[3]{RR}$) and Bazett's formula for QTcB (QTcB= QT/ $\sqrt[2]{RR}$) with RR=60/HR. Baseline QTcF and QTcB will be derived from the visit that other ECG parameters are flagged as baseline.

Listings of 12-lead ECGs will be provided with all relevant information and derived variables.

Unscheduled ECG measurements will not be used in computing the descriptive statistics for change from baseline at each post-baseline time point. However, they will be used in the analysis of notable ECG changes.

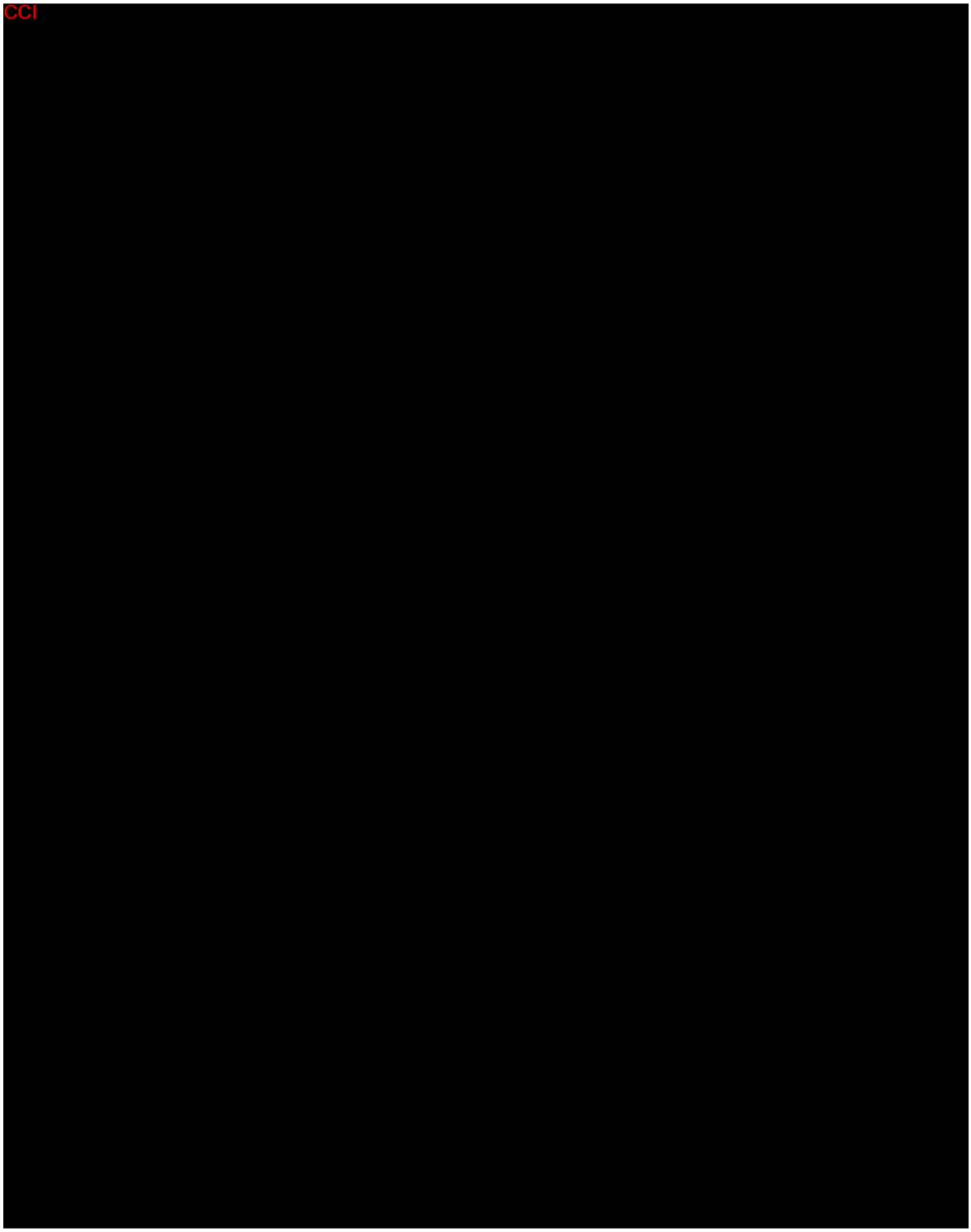
16.5.2 ECOG Performance Status

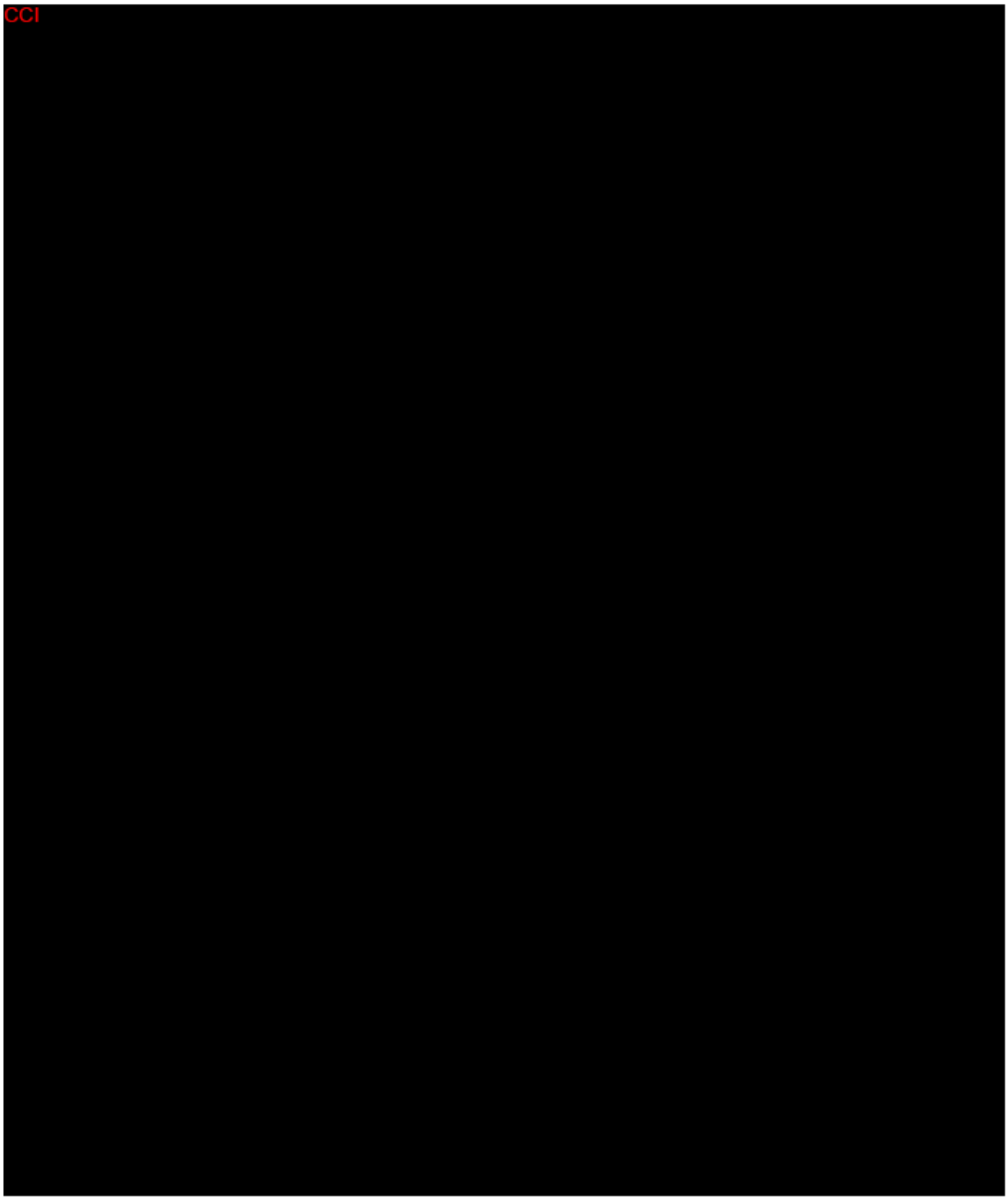
Analysis sets: Safety analysis set

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

CCI [redacted]

CCI




CCI

17 References

1. Hommel G, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. Statistics in Medicine 2007; 26(22): 4063-4073.
2. Jennison C, Turnbull BW. In: Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000.
3. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29-41.
4. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons 1980.
5. Cochran WG. Some Methods for Strengthening the Common Chi-square Tests. Biometrics 1954;10:417-51.
6. Agresti A. A Survey of Exact Inference for Contingency Tables. Statistical Science 1992;7:131–77.
7. Agresti A. An Introduction to Categorical Data Analysis (2 ed.). New York: John Wiley & Sons, 2007.
8. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. Journal of clinical oncology. 1998 Jan 1;16(1):139-44
9. Miettinen O, Nurminen M. Comparative analysis of two rates. Statistics in Medicine 1985;4(2):213-26
10. Liu GF, Wang J, Liu K, Snavely DB. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Statistics in Medicine 2006;30;25(8):1275-86

18 Appendices

Appendix I RECIST 1.1

Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-47.

Appendix II Clinically Important Protocol Deviations

Protocol Deviation	Protocol Deviation Code	Clinically Important PD	Proposed Check
Inclusion / Exclusion Criteria			
Subject did not meet inclusion criterion 4 (*Tumor determined to be positive for PD-L1 expression per the evaluation of a central laboratory*)	INCEXC01	y	Medical review required and Programming check on inclusion criterion
Subject did not meet inclusion criterion 6 (*Subjects with histologically confirmed metastatic or recurrent NSCLC (per 7th International Association for the Study of Lung Cancer classification)*)	INCEXC03	y	Medical review required and Programming check on inclusion criterion
Subject did not meet inclusion criterion 7 (*Subjects must not have received any treatment for systemic lung cancer, including EGFR inhibitors or anaplastic lymphoma kinase (ALK) inhibitors*)	INCEXC05	y	Medical review required and Programming check on inclusion criterion
Subject met exclusion criterion 1 or 2 (*Subjects whose disease harbors an activating EGFR mutation, or Subjects with non-squamous cell NSCLC whose disease harbors an ALK rearrangement*)	INCEXC06	y	Medical review required and Programming check on exclusion criterion
Subject met exclusion criterion 4 (*Prior therapy with any antibody or drug targeting T-cell coregulatory proteins, concurrent anticancer treatment, or immunosuppressive agents*)	INCEXC08	y	Medical review required and Programming check on exclusion criterion
Subject met exclusion criterion 5 (*Concurrent anticancer treatment (for example, cytoreductive therapy, radiotherapy [with the exception of palliative bone-directed radiotherapy], immune therapy, or cytokine therapy except for erythropoietin)*)	INCEXC09	y	Medical review required and Programming check on exclusion criterion
Subject met exclusion criterion 6 (*Major surgery for any reason, except diagnostic biopsy, within 4 weeks of randomization and / or if the subject has not fully recovered from the surgery within 4 weeks of randomization)*)	INCEXC10*	y*	Medical review required and Programming check on exclusion criterion
Subject met exclusion criterion 16 (*Pregnancy or lactation*)	INCEXC14	y	Medical review required and Programming check on exclusion criterion
Concomitant Medications/Procedures			
For avelumab arms, subjects took immunotherapy, immunosuppressive drugs (that is, systemic corticosteroids except for cases specified in a protocol section 6.5.2)	CONMED02	y	Medical review required
Subject had any anti-cancer treatment during the study (surgery, radiotherapy -unless bone directed or radiotherapy to a superficial lesion as long as there is no identified soft tissue component which has been selected as target lesion).	CONMED07	y	Medical review required
Protocol Procedures			

Protocol Deviation	Protocol Deviation Code	Clinically Important PD	Proposed Check
Study procedure not done/omitted: Screening CT scans not performed (lung/abdomen/pelvis, 6 weeks brain, bone scan if clincially indicated)	PROCED13	y	Medical review required and Programming check on CT scan recorded at screening
Randomization			
Subject randomised in error or randomisation interupted in IWRS resulting in erroneous application of the randomisation code	RANDOM01	y	
Patient randomized more than once	RANDOM02	y	Medical review required and Programming check on potential several subjects with the same date of birth, gender, race and baseline height

* This might be, but not necessarily is, a CIPD. This will have to be decided on a case-by-case basis during the data review meetings.

Appendix III EORTC QLQ-C30 version 3.0

Table 1: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range*	Version 3.0 Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL**:

$$Score = \{(RS-1)/range\} \times 100$$

Examples:

Emotional functioning

$$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$$
$$\text{EF } Score = \{1 - (RawScore - 1)/3\} \times 100$$

Fatigue

$$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$$
$$\text{FA } Score = \{(RawScore - 1)/3\} \times 100$$

Appendix IV EORTC QLQ-C30-LC13: Lung Cancer Module

Scoring of the lung cancer module

The lung cancer module incorporates one multi-item scale to assess dyspnoea, and a series of single items assessing pain, coughing, sore mouth, dysphagia, peripheral neuropathy, alopecia, and haemoptysis.

The scoring approach for the QLQ-LC13 is identical in principle to that for the symptom scales / single items of the QLQ-C30.

Scale name	Scale	Number of items	Item range*	QLQ-LC13 Item numbers	†
Symptom scales / items					
Dyspnoea†	LCDY	3†	3	3,4,5	X
Coughing	LCCO	1	3	1	
Haemoptysis	LCHA	1	3	2	
Sore mouth	LCSM	1	3	6	
Dysphagia	LCDS	1	3	7	
Peripheral neuropathy	LCPN	1	3	8	
Alopecia	LCHR	1	3	9	
Pain in chest	LCPC	1	3	10	
Pain in arm or shoulder	LCPA	1	3	11	
Pain in other parts	LCPO	1	3	12	

* "Item range" is the difference between the possible maximum and the minimum response to individual items.

† The dyspnoea scale should only be used if all three items have been answered. Some respondents ignore question 5 because they never climb stairs; in this case, the score for the dyspnoea scale would be biased if it were based upon the other two items. Hence if item 5 is missing then items 3 and 4 should be used as single-item measures.

Appendix V Description of the Case Definition for Assessment of Immune-Related AEs and IRRs

Immune-Related Adverse Events

Immune-related adverse events (irAEs) will be identified programmatically. AEs which satisfy all of the following criteria will be flagged as immune-related:

1) The AE preferred term matches a preferred term on the list of pre-selected MedDRA terms. The list of term is reviewed and updated, if necessary, for each new version of the MedDRA dictionary and the complete list of terms is included in a "lookup" dataset that is inlcuded in the analysis datasets for the study. Therefore, the list is not provided in this document.

2) The AE onset occurs after the first study drug administration and no more than 90 days after last dose.

3) On the AE eCRF page, the question "Were Corticosteroids, Immunosuppresants, or hormonal therapy (e.g. Thyroid) applied?" has the answer "Yes" selected.

4) On the AE eCRF page, either:

a. The question "Does any of the following provide a clear etiology for the event?" has the answer "No" selected, indicating that the AE is not attributable to underlying cancer disease/PD, prior or concomitant medications/procedures, nor another medical condition such as an infection or pre-existing disease.

OR

b. The AE eCRF indicates that a biopsy was performed and the question "Is the histopathology/biopsy consistent with an immune-mediated event?" has the answer "Yes" selected.

In the case that criteria (1) through (3) are met, and entries for condition (4) are missing, the following rules will be applied:

- If the answer to "Does any of the following provide a clear etiology for the event?" (4a) is missing, the event will be considered as irAE (irrespective of biopsy results).
- If the answer to "Is the histopathology/biopsy consistent with an immune-mediated event?" (4b) is missing, or if no biopsy was performed, and condition (4a) is not satisfied (i.e. "Yes" is selected as the answer to the question "Does any of the following provide a clear etiology for the event?"), the event will be considered as a non-irAE.

IRRs

Infusion related reactions are identified based on a list of MedDRA PTs and criteria on the timely relationship according to **Table 16**.

Table 16 Criteria for infusion related reactions

Infusion related reactions	**Reactions - Considered when onset is on the day of study drug infusion (during or after the infusion) or the day after the study drug infusion (irrespective of resolution date):** • Infusion related reaction • Drug hypersensitivity • Anaphylactic reaction • Hypersensitivity • Type 1 hypersensitivity **Signs and Symptoms - occurring on the day of study drug infusion (during or after the infusion) and resolved on the day of onset or the next day** • Pyrexia • Chills • Flushing • Hypotension • Dyspnoea • Wheezing • Back pain • Abdominal pain • Urticaria

Signature Page for VV-CLIN-280014 v2.0

Approval	PPD Clinical 11-Nov-2021 17:34:55 GMT+0000

Signature Page for VV-CLIN-280014 v2.0