Statistical Analysis Plan for Analysis of Efficacy and Safety

Clinical Trial Protocol Identification No.	EMR 100070-004
Title:	A Phase III open-label, multicenter trial of avelumab (MSB0010718C) versus docetaxel in subjects with non-small cell lung cancer that has progressed after a platinum-containing doublet
Trial Phase	Phase III
Clinical Trial Protocol Date and Version	10 January 2017/ Version 6.0
Statistical Analysis Plan Author	PPD , Statistical Scientist
Statistical Analysis Plan Date and Version	11 December 2017 /version 4.0
Statistical Analysis Plan Reviewers	PPD , Senior Biostatistics Reviewer
	PPD , Program Indication Lead
	PPD Medical Lead
	PPD PPD
	PPD , Biostatistics Program Lead
	PPD , QPD NBE Program Representative
	PPD PPD
	PPD , PPD
	PPD , Pathology
	PPD Senior Principal Statistical Programmer

A Phase III open-label, multicenter trial of avelumab versus docetaxel in subjects with non-small cell lung cancer that has progressed after a platinum-containing doublet

PPD

Statistical Scientist PPD

PPD

PPD Signature

Trial Biostatistician

PPD

PPD   Date of Signature   PPD

PPD
PPD

PPD

Signature
Project Statistician
Merck KGaA
Frankfurter Str. 250
64293 Darmstadt, Germany

PPD

PPD

PPD

PPD

Signature   PPD
Medical Responsible
EMD Serono 45A Middlesex Turnpike Billerica, MA 01820

2   Table of Contents

3   List of Abbreviations and Definition of Terms

ADR	Adverse Drug Reaction
CCI	████████████
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
CI	Confidence Interval(s)
CIPD	Clinically Important Protocol Deviations
CMH	Cochran-Mantel-Haenszel
CPI	Checkpoint Inhibitor
CR	Complete Response
CRF	Case Report Form
CSR	Clinical Study Report
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DD	Definitive Deterioration
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
EEA	European Economic Area
eCDF	Empirical Cumulative Distribution Function
eCRF	Electronic Case Report Form
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EGFR	Epidermal Growth Factor Receptor
EORTC	European Organization for Research and Treatment of Cancer

EQ-5D	EuroQOL Five Dimensions Questionnaire
FAS	Full Analysis Set
GCP	Good Clinical Practice
GGT	Gamma-glutamyl Transferase
HR	Hazard Ratio
HRQoL	Health-related Quality of Life
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
IEC	Independent Ethics Committee
IERC	Independent Endpoint Review Committee
IMP	Investigational Medicinal Product
INR	International Normalized Ratio
IPD	Important Protocol Deviation
irAE	Immune-related Adverse Event
IRR	Infusion-related Reaction
ITT	Intent-to-Treat
IUO	Investigation Use Only
IV	Intravenous
IWRS	Interactive Web Response System
LDH	Lactate Dehydrogenase
LLN	Lower Limit of Normal
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MedDRA	Medical Dictionary for Regulatory Activities
MID	Minimum Important Difference
MMRM	Mixed-effect Model Repeated Measures
MRI	Magnetic Resonance Imaging
nAb	Neutralizing Antibody
NCI	National Cancer Institute
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events

NE	Non-evaluable
NK	Natural Killer
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
CCI	CCI
PR	Partial Response
PRF	Pathology Report Form
RBC	Red Blood Cell
RECIST 1.1	Response Evaluation Criteria in Solid Tumors version 1.1
SAE(s)	Serious Adverse Event(s)
SAP	Statistical Analysis Plan
SD	Stable Disease
TEAE	Treatment-emergent Adverse Event
TUDD	Time until Definitive Deterioration
ULN	Upper Limit of Normal
VAS	Visual Analogue Scale

4   Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	13 February 2015	PPD	NA. The first version
2.0	28 February 2017		Details of changes are specified in Section 5.1 "Changes to Previous Version"
3.0	28 June 2017		Details of changes are specified in Section 5.1 "Changes to Previous Version"
4.0	11 December 2017		Details of changes are specified in Section 5.1 "Changes to Previous Version"

5   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 EMR 100070-004. 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 will 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 protocol amendments and is prepared in compliance with International Conference on Harmonization (ICH) E9. The first version (version 1.0) focuses on the detailed description of the primary and key secondary endpoints analysis and key safety endpoints analysis. Version 2.0 and version 3.0 were created based on the protocol version 5.0 dated 29 November 2016. Additional updates in version 2.0 and v3.0 are specified in Section 5.1, Changes to Previous Version.

5.1   Changes to Previous Version

Version 4.0

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

1. Section 10 *Analysis Set*: Subgroup Analysis set was updated for the PD-L1 subgroup variables. The original pathology scoring results were substituted by the rescoring results with a verified IUO (Investigational Use Only) assay for the determination of the PD-L1 expression status at 50% and 80% cut-off value. A new subgroup of "Recruitment waves" was added;

2. Section 13.3.3 *PD-L1 Expression Status and Baseline Biomarkers* was updated to include more details on PD-L1 rescoring at 50% and 80% cut-off values;

3. Section 16.1.2 *Sensitivity Analyses of Primary Endpoint* was updated to include a new sensitivity analysis of OS based on Safety analysis set as treated;

4. Section 16.1.3 *Subgroup Analysis of Primary Endpoint* was updated to include overall HR and 95% CI in the forest plot; the text was corrected to specify the point estimate of the interaction model parameter would be provided;

5. Section 17.1.1 *All Adverse Events* was updated to drop summary of related irAE;

6. Section 17.2.3 *Other Significant Adverse Events* was updated by adding details on two-level irAE summary; related irAE were deleted from the irAE analysis.

Version 3.0

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

1. Section 10 *Analysis Set* was updated for the definition of Safety Analysis Set; additional details were included for Subgroup Analysis Sets;

2. Section 11 *General Specifications for Statistical Analyses* was updated to include more details on baseline definition, summary statistics over time, imputation rules for AE stop date and date of subsequent anti-cancer therapy;

3. Section 13.1 *Demographics* was updated to add EEA;

4. Section 13.3.2 *Prior Anti-Cancer Therapies* was updated to include number of prior anti-cancer therapy regimens for metastatic or locally advanced disease;

5. Section 16.2.3 *Health-Related Quality of Life* was updated to include details on analysis summary of scheduled and unscheduled visits;

6. Section 16.2.4 *Sensitivity Analyses of Secondary Efficacy Endpoints* was updated to include additional sensitivity analysis for PFS per IERC and confirmed BOR per IERC;

7. Section 16.2.5 *Time of Follow-up for PFS* was added;

8. Section 16.2.6 *Subgroup Analysis of Secondary Efficacy Endpoints* was updated to remove CCI and subgroup analysis on HRQoL assessment. The details of HRQoL subgroup analysis will be included in the HRQoL supplemental SAP;

9. Section 17.1 *Adverse Events* was updated to including details of pooling the same AE with different toxicity grade, outcome or seriousness recorded as different entries on the eCRF;

10. Section 17.2.4 CCI

11. Section 17.5.3 CCI

12. Appendix V was updated to include detailed information for IRR.

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 or PD-L1+ PP analysis will be performed on this study per protocol version 4.0;

3. Section 7.1 Sample Size was updated per protocol version 4.0 (15 October 2015) for the total number of randomized subjects to be 750, based on an estimated prevalence of approximately 70% PD-L1+ subjects;

4. Section 7.2 Randomization was updated to drop randomization stratum of EGFR status per protocol version 3.0 (10 July 2015);

5. Section 8 Overview of Planned Analyses was updated to include O'Brien-Fleming boundaries based on a Lan-DeMets alpha spending function that account for the actual number of PD-L1+ events at the interim and the final analysis; data cut-off date for the interim and the final analysis was updated as pre-specified date;

6. Section 10 HRQoL analysis set and PD-L1+ HRQoL analysis set were added; Subgroup Analysis Sets was updated to include subgroup of PD-L1 tumor cell expression, age group 2, number of prior anti-cancer drug therapy and Central Nervous System (CNS) metastasis at baseline; reference was updated for some other subgroups;

7. Section 11 General Specifications for Statistical Analyses was updated to be consistent with master SAP template; visit windowing was added for laboratory assessment;

8. Section 12.1 Disposition of Subjects and Discontinuations was updated;

9. Section 12.2 Protocol Deviations was updated with current standard text; clinically important protocol deviations (CIPIDs) were added;

10. Section 13.1 Demographics – region classification was updated; age group of < 75 and >= 75 years old was added;

11. CCI ▓▓▓
▓▓▓▓▓▓

12. Section 14.1 Prior and Concomitant Medications/Procedures, definition of prior medications was updated and sorting order of the summary output was updated;

13. Section 14.2 Subsequent Anti-Cancer Therapies/Procedures was updated with more details;

14. Section 16.1.1 Primary Efficacy Analysis of Overall Survival was updated to include summary of events/censoring reasons;

15. Section 16.1.2 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;

16. Section 16.2.1 Best Overall Response was updated to clarify tumor assessment after start of any further anti-cancer will not be included in BOR derivation; Table 5 was updated to cite the original BOR derivation table from Eisenhauser, et al.; summary table was added for BOR of Non-evaluable (NE);

17. Section 16.2.2. Progression-Free Survival was updated to include additional censoring rules per master SAP template; summary of events/censoring reasons for PFS was added;

18. Section 16.2.3 Health Related Quality of Life was updated to include more details on the QoL scoring system; compliance, time until definitive deterioration (TUDD) analysis, and additional summary analysis including figures were added;

19. Section 16.2.4 Sensitivity Analyses of Secondary Efficacy Endpoints was updated;

20. CCI ███████████████████████████████████

21. CCI ███████████████████████████████████████

22. Section 17.1 Adverse Events, definition of immune-related AE and infusion-related reactions was updated;

23. Section 17.1.1 All Adverse Events, additional categories were added to the overall AE summary;

24. Section 17.2.3 Adverse Events of Special Interest was deleted; Other Significant Adverse Events was added to include Immune-related adverse event (irAEs) and Infusion-related reactions (IRRs);

25. Section 17.2.4 Immunogenicity Subgroup Analysis of Adverse Events was added;

26. Section 17.3.1 Hematology and Chemistry Parameters was updated;

27. Appendix II was updated to include clinically important protocol deviations by programming check and medical review;

28. 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.

29. New Appendix III was added to include the screenshot of EORTC QLQ-C30 version 3.0 scoring system;

30. Appendix IV was added to include the screenshot of EORTC QLQ-C30-LC13 Lung Cancer Module scoring system;

31. Appendix V was added to include "Description of the Case Review for Assessment of Immune-Related AEs".

6   Summary of Clinical Trial Features

Trial objectives	**Primary objective**
	To demonstrate superiority with regard to overall survival (OS) of

avelumab versus docetaxel in subjects with programmed death ligand 1 (PD-L1) positive (+; as determined by a companion diagnostic test under development), non-small cell lung cancer (NSCLC) after failure of a platinum-based doublet

Secondary objectives

Secondary objectives are as follows:

- To demonstrate superiority with regard to OS of avelumab versus docetaxel in the Full Analysis Set (FAS)

- To demonstrate superiority with regard to the objective response rate (ORR) of avelumab versus docetaxel in PD-L1+ subjects

- To demonstrate superiority with regard to progression free survival (PFS) of avelumab versus docetaxel in PD-L1+ subjects

- To demonstrate superiority with regard to the ORR of avelumab versus docetaxel in the FAS

- To demonstrate superiority with regard to PFS of avelumab versus docetaxel in the FAS

- To compare the subject-reported outcomes / quality of life when treated with avelumab versus docetaxel using the EuroQOL 5-dimensions questionnaire (EQ-5D) and the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 and module QLQ-LC13 in the FAS

- To determine the safety and tolerability of avelumab

Exploratory objectives

Exploratory objectives are as follows:

	CCI
Trial design and plan	This is a multicenter, international, randomized, open-label, Phase III trial in subjects with locally advanced unresectable, metastatic, or recurrent NSCLC that has progressed after a platinum doublet.
	Approximately 750 subjects, among them 522 PD-L1 assay positive subjects, will be randomized in a 1:1 ratio to receive either
	• avelumab at a dose of 10 mg/kg as a 1-hour intravenous (IV) infusion once every 2 weeks, or
	• docetaxel at a starting dose of 75 mg/m^2 (per label) by IV infusion once every 3 weeks.
	Subjects will be stratified according to PD-L1 assay status (positive versus negative expression in tumor cells) and NSCLC histology (squamous cell versus non-squamous cell).
	Tumor measurements by computed tomography (CT) scan or magnetic resonance imaging (MRI) will be performed every 6 weeks from randomization to determine response to treatment. A central imaging laboratory will be used to read and interpret all CT / MRI data; however, treatment decisions will be made by the treating Investigator. Response will be evaluated using the Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1).
	Treatment will continue until disease progression, significant clinical deterioration, unacceptable toxicity, any criterion for withdrawal from the trial or trial drug is fulfilled. For subjects receiving avelumab, treatment may continue past the initial determination of disease progression per RECIST 1.1 if the subject's performance status has remained stable, and if in the opinion of the Investigator, the subject will benefit from continued treatment and if other criteria are fulfilled as outlined in the protocol. Subjects receiving avelumab who have experienced a confirmed complete response (CR) should be treated for a maximum of 24 months after confirmation, at the discretion of the Investigator. If the Investigator believes that a subject may benefit from treatment beyond 24 months, it may be permissible after discussion with the Sponsor. In case a subject with a confirmed CR relapses after stopping treatment, but prior to the end of the trial, 1 re-initiation of treatment is allowed at the discretion of the Investigator and agreement of the Medical Monitor. In order to be eligible for re-treatment, the subject must not have experienced any toxicity that led to treatment discontinuation of the initial avelumab therapy. Subjects who re-initiate treatment will stay on trial and will be treated and monitored according to the protocol and the "until progression" schedule in the Schedule of Assessments.
	Patients assigned to docetaxel will not be allowed to crossover to avelumab as long as superiority of avelumab versus docetaxel in terms

	of the primary objective has not been demonstrated in the planned interim or final analysis.

Decisions regarding medical management of subjects will be made by the Investigator; however, the secondary endpoint determinations (response and disease progression) will be according to the central imaging assessment and review by a blinded Independent Endpoint Review Committee (IERC).

Adverse events (AEs) will be assessed throughout 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, the validity and scientific merit of the trial, and to evaluate efficacy at the 75% interim analysis. |
| **Planned number of subjects** | Approximately 750 subjects will be randomized. Accrual will proceed up to a target number of 522 PD-L1+ subjects. |
| **Primary endpoint** | The primary endpoint for the trial is OS time, defined as the time (in months) from randomization to the date of death. |
| **Secondary/Exploratory endpoints** | Secondary endpoints include

- PFS time according to RECIST 1.1 and as adjudicated by the IERC,
- Best overall response (BOR) according to RECIST 1.1 and as adjudicated by the IERC,
- changes in subject-reported outcomes / quality of life as assessed by the EQ-5D and the EORTC QLQ-C30 and module QLQ-LC13 questionnaire, and
- the safety profile of the trial drugs as measured by the incidence of AEs, SAEs, clinical laboratory assessments, vital signs, physical examination, ECG parameters, and ECOG PS.

Exploratory endpoints include

CCI |

	• CCI ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮
	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮
	▮▮▮▮▮▮▮▮▮▮▮▮▮

7   Sample Size/Randomization

7.1   Sample Size

The study is planned with a group sequential design, which includes an interim assessment for efficacy using an O'Brien-Fleming stopping boundary. The following assumptions with regard to the primary endpoint of OS time in the PD-L1+ subjects are made for the sample size calculation:

- Hazard ratio (HR) of 0.70 corresponding to an increase in median OS time from 8 months in the control arm to 11.43 months in the investigational arm under the exponential model assumption

- 1:1 randomization

- Alpha = 0.025 (1-sided)

- Power = 90%

- Uniform accrual over a period of 10 months

- A follow-up time of 11 months after randomization of the last subject

- An expected drop-out rate of 5%

- An interim analysis for efficacy after 75% of the planned events (deaths) in PD-L1+ subjects have been observed, with O'Brien-Fleming stopping boundary

A sample size of approximately 522 PD-L1+ subjects is planned in order to observe at least 337 events (deaths) at the final analysis. Calculations were performed using ADDPLAN v6.01, Aptiv Solutions. The 75% interim analysis will be performed after approximately 253 events are observed in PD-L1+ subjects. The actual data cut-off date for the interim analysis and the final analysis will be prospectively determined based on the event projection on PD-L1+ events provided by the unblinded team which is designated by the sponsor and is separate from the trial team. The actual number of PD-L1+ events may be slightly different from 253 for the interim analysis and 337 for the final analysis, respectively. An IDMC (see Section 2.3.1 of the clinical trial protocol) will be convened to perform the evaluation at the interim analysis in order to safeguard the Sponsor's personnel from unblinded trial results.

To ensure that the overall full analysis set (FAS) is representative of the underlying patient population, the study will enroll PD-L1 evaluable subjects without predetermining the ratio between PD-L1+ and PD-L1– subjects. Thus the total number of subjects enrolled in the study is an approximate estimate derived from prevalence information of PD-L1+ tumors from the NSCLC expansion cohort of Study EMR100070-001, which was conducted in a similar patient population as the current Study EMR100070-004. The preliminary estimate (at the time of initiation of current study EMR100070-004) of the proportion of PD-L1+ subjects, based on an analysis of CCI

CCI   in study EMR100070-001, was CCI
CCI   however, as the rate observed in the current study was later observed to be approximately 70%; the enrollment number was increased to approximately 750 subjects to target the planned 522 PD-L1+ subjects as per protocol version 6.0 (dated 10 January 2017).

7.2   Randomization

Qualified subjects will be randomized at a 1:1 ratio to receive either avelumab or docetaxel using stratified permuted block randomization with variable block length via the interactive web response system (IWRS). Randomization will be stratified according to PD-L1 assay status (positive versus negative expression in tumor cells) and NSCLC histology (squamous cell versus non-squamous cell). The purpose of stratification is to ensure balanced distribution of prognostic factors between treatment arms. Randomization will occur upon completion of the screening procedures and determination of subject eligibility, using the IWRS as described in Section 6.3 of the clinical study protocol.

The protocol version 3.0 (dated 10 July 2015) excluded further enrollment of NSCLC subjects with EGFR mutations. Therefore EGFR status is not considered as a randomization stratum for subjects enrolled under protocol version 3.0 and above. For analysis purposes, the subjects in the previously used strata non-squamous cell EGFR normal and non-squamous cell EGFR activating mutations will be combined into the stratum non-squamous cell.

8   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 as well as data of tumor assessment results as determined by the Independent Endpoint Review Committee (IERC), Health-related quality of life (HRQoL) data and biomarker data, which are collected by external vendors. 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 monthly event projection provided by the unblinded team as specified in Section 7.1. For the 75% interim analysis, the clinical cut-off date will be the date on which approximately 253 events (deaths) are expected in PD-L1+ subjects based on the event projections performed by the unblinded team. For the final analysis, the clinical cut-off date will be the date when approximately 337 events (deaths) in PD-L1+ subjects are expected based on the event projections performed by the unblinded team.

Since the observed number of events at the interim analysis may not be exactly equal to the planned 253 OS events, 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 PD-L1+ events deviates from the target number of 337 PD-L1+ events, 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 1, Section 8.2 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 OS, a statistically significant finding at the interim will be intended to claim superiority.

A separate SAP will cover the interim analysis for periodic safety review and interim efficacy review by the Independent Data Monitoring Committee (IDMC).

Separate or supplemental analysis plans might be written to cover:

- CCI

- Additional analyses of patient reported outcomes
- CCI

8.1   Sequence of Analyses

The following analyses will be performed during this trial. As the data cut-off date for both the interim efficacy and the final analysis will be prospectively determined based on event projection provided by the unblinded team, the actual number of PD-L1+ events may be slightly different from the planned number as indicated below.

- Interim efficacy analysis: will be performed after approximately 75% of events (deaths) required for final analysis have been observed in PD-L1+ subjects.
- Final analysis: will be performed after approximately 337 events have been observed in PD-L1+ subjects.

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

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

8.2   Interim Analysis

An IDMC will be formed and will be responsible for periodic safety evaluations of the trial as well as the evaluation of the 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 will provide details about the conduct of the IDMC meeting and decision-making rules.

For this trial there will be one interim efficacy analysis, which will be conducted after the prospectively determined data cut-off date when the number of PD-L1+ events reaches close to 253 based on the monthly event projection provided by the unblinded team. 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 1 for different number of PD-L1+ events at the interim analysis. The actual alpha spent at the interim analysis will be calculated based on the actual number of PD-L1+ events as of the data cut-off date for the internal analysis. The type I error and the efficacy

boundary for the final analysis will be adjusted accordingly based on the actual alpha spent at the interim analysis and the actual number of PD-L1+ events at the final analysis.

Table 1: Planned Lan-DeMets (O'Brien-Fleming) Efficacy Boundaries of OS

Look	Info Fraction (t)	Approx. Cum. Events	Cumulative Type I Error (1-sided)	Efficacy Boundaries z-scale
Interim	~75%	253	.00965	2.34
	76.5%	258	.01039	2.31
	78%	263	.01115	2.29
	79.5%	268	.01194	2.26
	81.0%	273	.01276	2.23
	82.5%	278	.01360	2.21
Final	100%	337	0.025	Will be adjusted based on interim outcome and the actual number of PD-L1+ events at the final analysis
Lan-DeMets alpha spending function for O'Brien-Fleming boundaries The overall type I error rate = 0.025 (1-sided) for primary endpoint of PD-L1+ OS. The exact efficacy boundaries may be updated prior to the time of analysis based on the actual number of events and the information fraction.				

The primary endpoint, OS, key secondary endpoints analysis including PFS and BOR, and safety analysis will be conducted in this interim analysis. The interim analysis of efficacy endpoints will be conducted on the FAS, with the primary analysis population being the PD-L1+ subset of the FAS. The IDMC will also be presented with subject disposition, subject background, baseline disease and demographic information, along with safety information.

After the prospectively determined number of PD-L1+ events are observed for the interim analysis, the independent statistical provider will prepare the outputs 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 this group to the IDMC only. The independent support IDMC statistician will be available at the IDMC meeting should any questions from the IDMC members arise regarding the data and / or analyses. Study staff involved with the day to day management of the trial, as well as any Sponsor staff, will not have access to this interim information until the trial is stopped at interim as recommended by the IDMC.

The hierarchical testing strategy will be employed for the interim analysis. If the one-sided p-value from the stratified log-rank test of OS is below the significance level at the interim analysis based on a Lan-DeMets alpha spending function for O'Brien-Fleming boundaries (See Table 1), the treatment difference will be claimed as statistically significant and the secondary endpoints will be tested within the hierarchical testing strategy at the same nominal significance level. Details of hierarchical testing strategy are provided in Section 16.

Since the formal efficacy boundaries will be used at the interim analysis for the statistical testing of OS, a statistically significant finding at the interim will be intended to claim superiority. If, at the time of the interim analysis, the OS of avelumab assigned subjects is shown to be superior to that of those randomized to the docetaxel group with a p-value based on the Lan-DeMets alpha spending function for O'Brien-Fleming boundaries (see Table 1), the IDMC may declare superior efficacy in the avelumab treatment arm compared with those randomized to receive docetaxel and recommend that the study be stopped early. In case of a positive interim analysis all analyses described in this SAP will be performed to facilitate CSR writing.

8.3   Final Analysis

All planned analyses outlined in this SAP will be performed for the final analysis or so called primary analysis. A partial database lock will be performed for both the interim and the final 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).

Subject follow-up for progression and survival will continue until 5 years after the last subject receives the last dose of avelumab. 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.

A data review meeting will be held prior to the interim analysis and the final analysis (for both partial and full database lock). In addition, no database can be locked and no randomization code should be unblinded and no subject's PDL1 status revealed until this SAP has been approved.

9   Changes to the Planned Analyses in the Clinical Trial Protocol

The statistical methods specified in this SAP are in accordance with protocol version 6.0 (dated 10 January 2017).

The data cut-off date for the interim and the final analysis will be prospectively determined based on the monthly event projection of number of PD-L1+ events provided by the unblinded team. Therefore, the actual number of PD-L1+ events at the interim and the final analysis may be slightly different from the planned number of events as specified in the protocol. The critical boundaries for the group sequential test will be derived from the predefined spending function as described in Section 8.2 and the type I error rate at the interim and the final analysis will be adjusted accordingly. This change to the planned analysis will ensure the overall one-sided significance level is preserved at 0.025 for primary endpoint of PD-L1+ OS.

The evaluation of anti-tumor activity in terms of PFS, BOR, and duration of response as provided by the IERC according to RECIST 1.1 will consider radiological assessments exclusively.

Additional sensitivity analysis will be performed including the assessment of the IERC per the oncologist.

10 Analysis Sets

Screening Analysis Set

The screening analysis set includes 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 this analysis set.

PD-L1+ FAS

The PD-L1+ FAS is a subset of the FAS including all PD-L1+ subjects as recorded in the IWRS.

The primary analysis population will be the PD-L1+ FAS.

Safety Analysis Set

The Safety analysis set will include all subjects who were administered at least one dose of the study medication, i.e. avelumab or docetaxel. Subjects will be classified according to the treatment assigned at randomization unless the incorrect treatment was received throughout the dosing period in which case the subjects will be classified according to the first study treatment received.

PD-L1+ Safety Analysis Set

The PD-L1+ Safety analysis set is a subset of the Safety Analysis Set including all PD-L1+ subjects.

HRQoL Analysis Set

The HRQoL analysis set is a subset of the FAS and includes all FAS subjects who meet all of the following criteria:

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

PD-L1+ HRQoL Analysis Set

The PD-L1+ HRQoL analysis set will include all PD-L1+ subjects according to IWRS who were included in the HRQoL analysis set.

CCI

CCI

Subject assignment per the IWRS randomization stratum of PD-L1 status will be used to determine the PD-L1+ subset. PD-L1 positivity is defined by at least 1% PD-L1 positive tumor cells, as determined by a companion diagnostic assay.

Table 2 summarizes the use of the analysis sets in the different analyses.

Table 2   Statistical Analysis by Analysis Set

Analyses	FAS	PD-L1+ FAS	HRQoL Analysis Set	PD-L1+ HRQoL Analysis Set	Safety Analysis Set	PD-L1+ Safety Analysis Set
Baseline Characteristics	✓	✓	✓	✓	✓	✓
Past and Concomitant Therapies	✓	✓				
Important and Clinically Important Protocol Deviations	✓	✓				
Compliance and Exposure					✓	✓
Efficacy: Primary	✓	✓				
Efficacy: Secondary – PFS and BOR	✓	✓				
Efficacy: Secondary – HRQoL			✓	✓		
CCI						
Safety					✓	✓

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 defines the reference always. 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 following definition.

CCI

In the case of a low number of subjects within a category (< 25 subjects, which is about 5% of the randomized PD-L1+ population), the categories may be pooled when meaningful. The subgroup analysis will not be performed on any subgroup category "Missing".

PD-L1 subgroups are only relevant to the FAS analyses. The PD-L1 expression status (Positive/ Negative/ Non-evaluable), at cut-off values of 1%, 50% and 80%, is based on the results per the

pathology review form. The PD-L1 expression status at 1% cut-off is based on the original scoring, and the PD-L1 expression status at 50% and 80% cut-off is based on the rescoring. The analyses are described with more details in Section 13.3.3 PD-L1 Expression Status and Baseline Biomarkers

The following subgroups will be defined:

- 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
- Ethnicity
  - Hispanic/Latino
  - Non-Hispanic/Latino (Reference)
- Ethnicity
  - Japanese living in Japan
  - All other subjects (Reference)
- 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

- Smoking status
  - Never smoking (Reference)
  - Ever smoker

- Central Nervous System (CNS) metastasis at baseline as collected on the "Tumor Assessment" eCRF page
  - Yes
  - No (Reference)

- Number of prior anti-cancer drug therapies
  - 1 Prior therapy (Reference)
  - $\geq$ 2 Prior therapies

- EGFR mutation status collected on the "Disease History" eCRF page
  - Normal (Reference)
  - Activating mutation
  - Other abnormality
  - Not applicable

- ALK mutation status as collected on the "Disease History" eCRF page
  - Normal (Reference)
  - Rearrangement
  - Other abnormality
  - Not applicable

- PD-L1 tumor cell expression at cut-off of 1% (only apply to FAS analysis)
  - PD-L1 expression < 1% and non-missing (Reference)
  - PD-L1 expression $\geq$ 1%

- PD-L1 tumor cell expression at cut-off of 50% (only apply to FAS analysis)
  - PD-L1 expression < 50% (Reference)
  - PD-L1 expression $\geq$ 50%

- PD-L1 tumor cell expression at cut-off of 80% (only apply to FAS analysis)

  - PD-L1 expression < 80% (Reference)

  - PD-L1 expression $\geq$ 80%

- PD-L1 total % tumor cells (only apply to FAS analysis)

  - Negative: $\geq$ 0% to < 1% (Reference)

  - Positive: $\geq$ 1% to <50% at any PD-L1 intensity

  - Positive: $\geq$ 50% to <80% at any PD-L1 intensity

  - Positive: $\geq$ 80% at any PD-L1 intensity

  - Non-evaluable

- Recruitment waves defined based on countries

  - Recruitment wave 1 (Early wave): Subjects enrollment began in the early phase of the study, i.e. during the year of 2015. These subjects had earlier post-study access to marketed checkpoint inhibitors and were mainly enrolled in South Africa, Japan, Taiwan, Australia, Israel, United States, Belgium, Denmark, France, Spain, Switzerland (Reference).

  - Recruitment wave 2 (Late wave): Subjects enrollment began in the later phase of the study, i.e. during the year of 2016. These subjects had later post-study access to marketed checkpoint inhibitors and were mainly enrolled in Korea, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Poland, Romania, Russian Federation, Argentina, Brazil, Chile, Colombia, Mexico, Peru, Turkey, Italy, United Kingdom.

11   General Specifications for Statistical Analyses

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

Baseline characteristics summary and the efficacy analysis will be performed on the PD-L1+ FAS and the FAS, with the primary analysis population being the PD-L1+ FAS. Analyses performed on the FAS will take into account subjects' allocation to treatment groups as randomized. Analyses performed on the safety population will consider subjects as treated.

Data handling after cut-off date:

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

Pooling of centers:

In order to provide overall estimates of treatment effects, data will be pooled across all 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.

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 16.1 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 exploratory. 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 exploratory 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, [i.e. n (missing)], mean, median, standard deviation (StDev), 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 study at that visit, unless otherwise specified.

Definition of baseline:

For the safety analysis, baseline is defined as the last measurement before the first dose of study treatment. Baseline for heart rate and QTc assessments will be derived from the visit where both heart rate and QT are not missing.

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

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 (EOT) visit). Data reported at the EOT visit are not eligible for baseline selection.

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

Definition of 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 docetaxel).

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.

The date of new anti-cancer drug therapy is derived as outlined in Section 14.2.

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, StDev, median, minimum, maximum and quartiles) by nominal visit or time point for safety endpoints such as laboratory measurements, ECGs and vital signs will include only data from scheduled visits per protocol**.**

Summary statistics over time:

For descriptive statistics over time by nominal visit or time point for QoL data and safety endpoints (laboratory, ECG and vital signs), only those planned 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.

Visit window for local laboratory assessment:

The assignment of visit window is described in Table 3 for the purpose of by-visit analyses of laboratory assessment data:

- Baseline will be derived as described above.

- Both scheduled and unscheduled assessments are included for visit windowing.

- No visit windowing will be performed at Discontinuation, End-of-Treatment, or Safety Follow-up visits for laboratory assessment. Instead, the earliest non-missing observation among the unscheduled or scheduled assessments for each visit (Discontinuation, End of Treatment, and Safety Follow-up) will be used for the analysis.

- If there are multiple assessments for any specified visit and some of them are from scheduled visits with non-missing assessment results, the assessment from scheduled visit that is closest to the planned study day will be used for analysis.

- If there are multiple assessments for any specified visit and none of them are from scheduled visits, the assessment with non-missing results and closest to the planned study day will be used for analysis.

- If there are two or more unscheduled assessments with non-missing results and the same distance to the planned study day, the assessment prior to the planned study day will be used in deriving visit window.

Table 3   Visit Window Definition for Laboratory Assessment

Avelumab treatment arm:

Assigned Study Day (Inclusive)		Planned Study Day (AWTARGET)	Analysis Visit (N) (AVISITN)	Analysis Visit (AVISIT)
From (AWLO)	To (AWHI)			
~	1		1	Baseline
1	11	1	2	Week 1 Day 1-11
12	25	15	3	Week 3 Day 12-25
26	39	29	5	Week 5 Day 26-39
40	53	43	7	Week 7 Day 40-53
54	67	57	9	Week 9 Day 54-67
68	81	71	11	Week 11 Day 68-81
82	95	85	13	Week 13 Day 82-95
96	109	99	15	Week 15 Day 96-109
110	123	113	17	Week 17 Day 110-123
124	137	127	19	Week 19 Day 124-137
138	151	141	21	Week 21 Day 138-151
152	165	155	23	Week 23 Day 152-165
166	179	169	25	Week 25 Day 166-179
180	193	183	27	Week 27 Day 180-193
194	207	197	29	Week 29 Day 194-207

208	221	211	31	Week 31 Day 208-221
222	235	225	33	Week 33 Day 222-235
236	249	239	35	Week 35 Day 236-249
250	263	253	37	Week 37 Day 250-263
264	277	267	39	Week 39 Day 264-277
278	291	281	41	Week 41 Day 278-291
292	305	295	43	Week 43 Day 292-305
306	319	309	45	Week 45 Day 306-319
320	333	323	47	Week 47 Day 320-333
334	347	337	49	Week 49 Day 334-347
348	361	351	51	Week 51 Day 348-361
362	375	365	53	Week 53 Day 362-375
376	389	379	55	Week 55 Day 376-389

Docetaxel treatment arm:

Assigned Study Day (Inclusive)		Planned Study Day (AWTARGET)	Analysis Visit (N) (AVISITN)	Analysis Visit (AVISIT)
From (AWLO)	To (AWHI)			
~	1		1	Baseline
1	18	1	2	Week 1 Day 1-18
19	39	22	4	Week 4 Day 19-39
40	60	43	7	Week 7 Day 40-60
61	81	64	10	Week 10 Day 61-81
82	102	85	13	Week 13 Day 82-102
103	123	106	16	Week 16 Day 103-123
124	144	127	19	Week 19 Day 124-144
145	165	148	22	Week 22 Day 145-165
166	186	169	25	Week 25 Day 166-186
187	207	190	28	Week 28 Day 187-207
208	228	211	31	Week 31 Day 208-228
229	249	232	34	Week 34 Day 229-249
250	270	253	37	Week 37 Day 250-270
271	291	274	40	Week 40 Day 271-291
292	312	295	43	Week 43 Day 292-312
313	333	316	46	Week 46 Day 313-333
334	354	337	49	Week 49 Day 334-354
355	375	358	52	Week 52 Day 355-375
376	396	379	55	Week 55 Day 376-396
397	417	400	58	Week 58 Day 397-417

Standard derivations and reporting conventions:

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

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

Demographics and physical measurements:

- Age [years]:

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

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

- BMI (kg/m^2) = weight (kg)/[height (m)]^2.

- BSA (m²) = ([height (cm) x weight (kg)] / 3600)$^{0.5}$

For reporting conventions, mean and median should generally to be displayed one more decimal place than the raw data and standard deviation should be displayed to two more decimal places than the raw data. Percentages will be reported to one decimal place. The rounding will be performed to closest integer / first decimal using the common mid-point between the two consecutive values. 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 re-initiated treatment:

Data collected after re-initiation of treatment will not be summarized for safety and efficacy analyses except for overall survival and disposition. 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.

Missing data and imputation rules:

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" or "NA". For example, if N=1, the measure of variability (StDev) cannot be computed and should be presented as "ND".

Partial dates will be imputed as follows:

Disease history

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.

Adverse events

Incomplete AE-related dates will be imputed as follows:

- If the AE onset date is missing completely, then the onset date will be replaced by the start of study treatment.

- If only the day part of the AE onset date is missing, but the month and year are equal to the start of study treatment, then the AE onset date will be replaced by the start of study treatment.

  For example, if the AE onset date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed AE onset date will be 15/JAN2015.

- If both the day and month of the AE onset date are missing but the onset year is equal to the start of study treatment, then the onset date will be replaced by the start of study treatment. For example, if AE onset date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed AE onset date will be 19/NOV/2014.

- In all other cases the missing onset day or missing onset month will be replaced by 1.

- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of 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 (such as fatal outcome) might be taken from Safety database and included separately into CSR.

Prior/concomitant medication

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

- If the medication date is missing completely, then the medication date will be replaced by the start of study treatment.

- If the day of medication date is missing, but the month and year are equal to the start of study treatment, then the medication date will be replaced by the start of study treatment.

For example, if the medication start date is --/JAN/2015, and study treatment start date is 15/JAN/2015, then the imputed medication start date will be 15/JAN2015.

- If both the day and month of medication start date are missing but the start year is equal to the start of study treatment, then the medication date will be replaced by the start of study treatment. For example, if the medication start date is --/---/2014, and study treatment start date is 19/NOV/2014, then the imputed medication start date will be 19/NOV/2014.

- In all other cases the missing medication day or missing medication month will be replaced by 1.

- Incomplete stop date will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of 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.

Subsequent anti-cancer therapy

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.

Exposure

- 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

Date of last contact

The last contact date will be the latest complete date among the following:

- All subject assessment dates (laboratory blood draws, 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" (only used if status is 'alive')

- Study drug start and end dates

- Randomization date

- Date of discontinuation from the "Study Termination" eCRF page (do not use if reason for discontinuation is lost to follow-up)

Death date

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

- If the date is missing it will be imputed as day after date of last contact from the eCRF survival page
- If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) date of last contact (excluding the date of death) 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

Tumor assessments

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

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.

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

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

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

All statistical analyses will be performed using SAS® Version 9.2 or higher, or R (www.r-project.org), version 3.2.5 or higher.

12   Trial Subjects

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

12.1   Disposition of Subjects and Discontinuations

The following will be summarized overall and by treatment group, where appropriate. The summary will be presented for all FAS subjects and PD-L1+ FAS subjects. The percentages below will be calculated based on the number of subjects in the FAS and PD-L1+ FAS, respectively.

- Total number of subjects screened overall

- Number of subjects who screened failed overall and grouped by the main reason

- Number and percentage of randomized subjects in the following analysis sets:

  - FAS

  - PD-L1+ FAS

  - Safety analysis set

  - PD-L1+ Safety analysis set

  - HRQoL analysis set

  - PD-L1+ HRQoL analysis set

- Number of subjects randomized but not treated

- Number of randomized subjects still on treatment

- Number of randomized subjects completed treatment

- Number of randomized subjects per reason of treatment discontinuation

- Number of subjects who discontinued the treatment but are still in follow-up

- Number of subjects who re-initiated avelumab treatment

  - Number of subjects discontinued the treatment after re-initiation

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

- Number of randomized subjects overall, by region (European Economic Area (EEA)) (required by EudraCT), North America, Latin America, Western Europe, Eastern Europe, Asia, Australasia, Africa and Middle East), by country within region

- Number of randomized subjects by center

- Number of randomized subjects by randomization strata (IWRS)

- Number of randomized subjects by randomization strata (eCRF)

- Cross tabulation: stratum by IWRS vs. stratum by eCRF. This summary will only include stratum NSCLC histology. For stratum of PD-L1 status, the result will be collected in an external dataset and will be provided by the IWRS vendor, which will then be integrated into the clinical database. Therefore, no cross tabulation of PD-L1 status will be summarized.

- Cross tabulation: subjects randomized (avelumab/docetaxel) vs. subjects treated (avelumab/docetaxel/no)

  - Reasons if randomized≠treated

12.2   Protocol Deviations

Analysis sets: FAS / PD-L1+ FAS

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, the following IPDs are identified as clinically important protocol deviations (CIPD) which are a subset of important protocol deviations:

- Subjects do not meet inclusion criterion #4, i.e. Tumor determined to be evaluable for PD-L1 expression per the evaluation of a central laboratory

- Subjects do not meet inclusion criterion #5, i.e. Subjects with histologically confirmed Stage IIIb/IV or recurrent NSCLC who have experienced disease progression

- Subjects do not meet inclusion criterion #6, i.e. Subjects must have progressed after platinum based treatment for metastatic disease (minimum 2 cycles of treatment), or progressed within 6 months of completion treatment for locally advanced disease

- Subjects meet exclusion criterion #2, i.e. Systemic anticancer treatment was administered after progression during or after platinum-based combination treatment

- Subjects meet exclusion criterion #3, i.e. Negative ALK rearrangement and EGFR mutation status (note, under protocol version 1+2, EGFR mutation positive patients are eligible if failed also prior TKI treatment)

- Subjects meet exclusion criterion #4, i.e. Subjects have received prior immunotherapy with checkpoint inhibitors such as anti-PDL1/ PD-1, anti CTLA-4

- Subjects meet exclusion criterion #5, i.e. Subject was found to be receiving concurrent anticancer treatment

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

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.

13 Demographics and Other Baseline Characteristics

Analysis sets: FAS / PD-L1+ FAS / Safety Analysis Set / PD-L1+ Safety Analysis Set / HRQoL Analysis Set / PD-L1+ HRQoL Analysis Set

13.1 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, $\geq$ 65 years
    - 65-74, 75-84, $\geq$ 85
  - Pooled Region:
    - US 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) (m²)

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, sex, race, height (cm), weight (kg), BMI (kg/m^2), BSA, and ECOG performance status.

13.2   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.

13.3   Other Baseline Characteristics

13.3.1   Disease Characteristics

Information on disease characteristics collected on "Disease History" eCRF page will be summarized in total, by treatment arm as well as by treatment arm within strata, if applicable. Summary statistics will be presented for:

- Site of primary tumor

- Time since documented, locally advanced, inoperable or metastatic disease (months), defined as (date of randomization – date of documented, locally advanced, inoperable or metastatic disease)/30.4375

- Tumor histopathologic / cytologic type
  - Squamous cell
  - Non-squamous cell
    - Adenocarcinoma
    - Bronchoalveolar
    - Large cell
    - Other

- TNM classification at initial diagnosis

- TNM classification at study entry

- PD-L1 assay status as per IWRS
  - Positive
  - Negative

- EGFR mutation status
  - Normal vs activating mutations

- ALK mutation status
  - Normal
  - Rearrangement
  - Other abnormality

- CNS metastasis at baseline
  - Yes
  - No

- 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

Baseline characteristics with respect to ECOG, vital signs, physical examinations, ECG and hematology/biochemistry will be part of Section 17 (Safety Evaluation).

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

13.3.2   Prior Anti-Cancer Therapies

The prior anti-cancer therapies are collected under the "Prior Anti-Cancer Drug Therapies Details", "Prior Anti-Cancer Radiotherapy Details" and "Prior Anti-Cancer Surgeries Details" eCRF pages.

The number 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 any prior anti-cancer therapy regimens: missing / 1 / 2 / 3 / $\geq$4
- Number of prior anti-cancer therapy regimens for metastatic or locally advanced disease: missing / 1 / 2 / 3 / $\geq$4
- Type of prior anti-cancer therapy:
    - Cytotoxic therapy / Endocrine therapy / Monoclonal antibodies therapy / Small molecules / Immunotherapy / Other
    - EGFR-targeting therapy / Other therapy
- Intent of Therapy: Neo-Adjuvant / Adjuvant / Metastatic or 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

Timing related to the first date of prior anti-cancer therapy for metastatic or locally advanced disease (before/ after) will be listed. If the sampling date is prior to the first date of prior anti-cancer therapy for metastatic or locally advanced disease, it is considered as 'before'; otherwise, it is considered as 'after'.

13.3.3   PD-L1 Expression Status and Baseline Biomarkers

PD-L1 expression status will be summarized for each tumor cell at 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 PRF. PD-L1 expression status at the 1% cut-off is from the original scoring, while the 50% and 80% cut-off values are from the rescoring. Results will be broadly summarized as "Positive", "Negative" and "Non-evaluable". PD-L1 expression status at 50% and 80% cut-off values is derived as "Negative" if the original scoring at 1% cut-off value is "Negative" per the PRF. The details are described in Table 4.

The % of tumor cells at each PD-L1 staining intensity (0, 1+, 2+, and 3+) is also collected on both the original and the rescoring PD-L1 PRFs. The results based on the total % of tumor cells at each PD-L1 staining intensity will be summarized based on the category described in Table 5 at 50% and 80% cut-off on both the original scoring and the rescoring.

Table 4   PD-L1 Expression Status per Pathology Assessment

Original Scoring: Cut-off value at 1%	Rescoring: Cut-off value at 50% and 80%
Negative	Negative
Positive	Positive
	Negative
	Non-evaluable
Non-evaluable	n/a

Table 5   PD-L1 Expression Status per Total % of Tumor Cells Categorization

Original Scoring: Cut-off value at 1%	Total % of tumor cells at each PD-L1 staining intensity (0, 1+, 2+, and 3+): Original Scoring & Rescoring
Negative	Negative: original scoring; Negative: $\geq 0\%$ to <1% at any PD-L1 staining intensity for rescoring
Positive	Positive: $\geq 1\%$ to <50% at any PD-L1 staining intensity
	Positive: $\geq 50\%$ to <80% at any PD-L1 staining intensity
	Positive: $\geq 80\%$ at any PD-L1 staining intensity

The listing of PD-L1 expression data will include all the information collected on the PRF for each tumor cell and immune cell.

14   Previous or Concomitant Medications/Procedures

Analysis sets: FAS / PD-L1+ FAS

14.1   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 during the on-treatment period as well as those started during on-treatment period. **Previous 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 previous) 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 preferred term. 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. Number 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.

14.2   Subsequent Anti-Cancer Therapies/Procedures

Anti-cancer treatment after discontinuation will be provided in a data listing with data retrieved from "Anti-Cancer Treatment after Discontinuation Details", "Radiotherapy after

Discontinuation", and "Surgery after Discontinuation" eCRF pages. 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 of subjects with any anti-cancer treatment after discontinuation will be tabulated overall and by type of therapy based on the data collected from the "Anti-Cancer Treatment after Discontinuation Details" eCRF page: Cytotoxic therapy / Endocrine therapy / Monoclonal antibodies therapy / Small molecules / Immunotherapy / Other. In addition, the number and percentage of subjects who received subsequent immune therapy such 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 therapy 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 "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 Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term. The same approach as prior and concomitant medications will be applied in presenting the summary table of subsequent anti-cancer treatment.

15 Treatment Compliance and Exposure

Analysis sets: Safety Analysis Set / PD-L1+ Safety Analysis Set

All dosing calculations and summaries will be based on "Avelumab Administration Details" and "Docetaxel Administration Details" eCRF pages.

Subjects randomized to the avelumab arm will receive an IV infusion of avelumab at a dose of 10 mg/kg over the duration of 1 hour once every 2 weeks (one cycle). Subjects randomized to the docetaxel arm will receive IV infusion of docetaxel (Hospira) according to its label at a starting dose of 75 mg/m^2 by IV infusion over the duration of 1 hour once every 3 weeks (one cycle).

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

Avelumab:

For subjects randomized to avelumab, 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 study 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

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

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.

Docetaxel:

For subjects randomized to docetaxel, the dose level is calculated as actual dose administered /derived BSA (mg/m^2). The last available weight of the subject on or prior to the day of dosing will be used in deriving BSA.

The duration of docetaxel treatment (in weeks) during the study is defined as:

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

The cumulative dose (mg/m^2) of docetaxel 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) / derived BSA (m^2)).

Each cycle is defined by a 3-week period. The dose intensity and the relative dose intensity of docetaxel will be calculated for each subject across all cycles. The dose intensity (mg/m^2/cycle) of docetaxel per cycle is defined as

$$\text{dose intensity} = \left(\frac{\text{Cumulative dose of docetaxel}}{(\text{duration of docetaxel (in weeks)})/3}\right)$$

The relative dose intensity of docetaxel is defined as the actual dose intensity divided by the planned dose as specified in the protocol per cycle.

The following summary tables will be provided for both treatment groups:

- Duration of therapy (weeks)
- Total number of infusions received
- Cumulative dose of therapy (mg/kg for avelumab; mg/m² for docetaxel)

- Dose intensity (mg/kg/2 weeks for avelumab; mg/m²/3 weeks for docetaxel)
- Overall relative dose intensity of therapy (%)

Dose Reduction

A dose reduction is defined as actual non-zero dose < 90% of the planned dose. Number of subjects with at least one dose reduction as well as a breakdown of dose reductions (1 / 2 / 3 / ≥4) will be summarized for both treatment groups.

Dose Delay

Delays will be derived based on infusion date for both 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 of subjects with delayed infusions and maximum length of delay, i.e. the worst case of delay if subjects have multiple dose delays will be summarized.

Avelumab Infusion Rate Reductions

Infusion rate reductions as recorded on the eCRF will be used for analysis. Number 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 group.

A listing of study drug administration will be created with the information collected on the "Avelumab Administration Details" and "Docetaxel Administration Details" eCRF page.

16　　　　Endpoint Evaluation

The primary endpoint of the study is the overall survival (OS), defined as the time (in months) from randomization to the date of death, regardless of the actual cause of the subject's death.

The key secondary efficacy endpoints include PFS and BOR according to RECIST 1.1 and as assessed by the IERC. Radiological assessments will be used including adjudicated radiological reviews. Additional sensitivity analysis will also be performed utilizing the final clinical review by the oncologist.

A hierarchical testing strategy will be applied to perform confirmatory analysis of primary endpoint and the key secondary efficacy endpoints at both the interim and the final analysis as follows:

1. OS in the PD-L1+ FAS

2. OS in the FAS

3. BOR in the PD-L1+ FAS

4. PFS in the PD-L1+ FAS

5. BOR in the FAS

6. PFS in the FAS

Statistical testing of each of the hypotheses 2 through 6 will be performed at the same nominal significance level as for hypothesis 1 at the interim and the final analysis based on a Lan-DeMets alpha spending function for O'Brien-Fleming boundaries (see Section 8.2) if, and only if, the preceding null hypothesis in terms of the hierarchical order was rejected. In case the hierarchical testing procedure stops at a certain level at the interim analysis, the remaining hypotheses will be tested at the primary analysis with the remaining significance level based on Lan-DeMets alpha spending function for O'Brien-Fleming boundaries to control the overall Type 1 error rate at 0.025 (one-sided).

P-values that are reported apart from the inferential testing procedure are for exploratory purpose only.

All sensitivity analyses are regarded as purely exploratory; therefore, no formal adjustment for multiplicity will be undertaken for these sensitivity analyses.

16.1   Primary Endpoint Analyses

16.1.1   Primary Efficacy Analysis of Overall Survival

Analysis sets: PD-L1+ FAS / FAS

The primary efficacy analysis of OS will be performed on PD-L1+ FAS and FAS in the given order based on the hierarchical testing strategy described in Section 16. If the one-sided p-value from the stratified log-rank test of the OS on PD-L1+ FAS is below the significance level based on the Lan-DeMets alpha spending function for O'Brien-Fleming boundaries (see Table 1, Section 8.2) at the interim and the final analysis, respectively, then the same analysis will be repeated on the FAS.

All data required for the calculation of time to event will be taken from the eCRF. Randomization strata will be taken as specified and documented in IWRS.

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 11 as of the data cut-off date for the analysis.

OS time (in months) = (Date of death – date of randomization +1) / 30.4375

The primary efficacy analysis will compare the OS time between the two treatment groups, 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 1, Section 8.2. For the primary efficacy analysis performed on PD-L1+ FAS, the stratification factor is NSCLC histology as captured via the IWRS at randomization for PD-L1+ subjects; for the primary efficacy analysis performed on FAS, the stratification factors are PD-L1 assay status and NSCLC histology as captured via the IWRS at randomization. For analysis purposes, the subjects in the previously used strata non-squamous cell EGFR normal and non-squamous cell EGFR activating mutations will be combined into the stratum non-squamous cell.

The following null hypothesis will be tested:

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

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

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), 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=docetaxel, 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 (1), will be used to construct the two-sided CI for the interim and the final analysis. The alpha value as specified for some example situations in Table 1, Section 8.2 will be applied to calculate the two-sided CI for the interim analysis and the final analysis, respectively.

In additional, the unadjusted 95% CI of the HR will also be calculated at the interim and the primary analysis in order to provide comparable results of the sensitivity analysis of the primary endpoint as specified in Section 16.1.2.

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 survival rate at 6 and 12 months will be estimated with corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (2) 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 (3) (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of 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 page prior to the analysis cut-off;
- Subjects with the last contact date > 14 weeks prior to the analysis cut-off date (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.

16.1.2　　　　Sensitivity Analyses of Primary Endpoint

Analysis sets: PD-L1+ FAS / FAS

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

- An unstratified analysis will be performed to compare the OS time and to estimate the treatment effect;
- The primary analysis will be repeated using strata according to eCRF instead of IWRS data;
- If the actual number of PD-L1+ events is 10% more than the planned number of PD-L1+ events at the interim analysis and the final analysis, i.e. if there are > 278 PD-L1+ events and > 371 PD-L1+ events based on the prospectively determined cut-off date at the interim analysis and the final analysis, respectively, the primary analysis will be repeated by including only the first 253 and 337 PD-L1+ events that were observed in the study for the interim analysis and the final analysis, respectively.
- The primary analysis will be repeated by censoring those FAS 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 immune therapy minus 1 day. The final list of subsequent immune therapy will be provided upon medical review of all subsequent anti-cancer therapies.

- The primary analysis will be repeated on the safety analysis set as treated.

Methods for evaluating the validity of model assumptions

The proportional hazards assumption will be checked visually for the primary analysis by plotting log(-log(OS)) 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 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 arm, largest observed survival time for docetaxel arm).
- τ_2 = minimum of (largest survival event time for avelumab arm, largest survival event time for docetaxel arm).
- τ_3 = midpoint between τ_1 and τ_2.

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

The treatment effect between the avelumab arm and the docetaxel 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 two treatment arms will be plotted against time τ.

Exploratory analyses to investigate the impact of potential prognostic or effect modifying (predictive) factors (see subgroup as given in Section 10)

Multivariable Cox regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. The Cox's Proportional Hazard model is defined as: $h(t) = h(0;t) \exp(Xb)$, 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 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% CIs. No interactions will be considered. Post-baseline factors will not be considered for the model.

16.1.3   Subgroup Analysis of Primary Endpoint

Analysis sets: PD-L1+ FAS / FAS

The subgroup analyses will be performed on the primary endpoint for all subgroup levels defined in Section 10 "Subgroup Analysis Sets". All the subgroup analyses are exploratory; no adjustment for multiplicity will be performed. In the case of a low number of subjects within a category (< 25 subjects, which is about 5% of the randomized PD-L1+ population), the categories may be pooled when meaningful.

All the subgroup analyses will be performed as unstratified.

The OS time between the two treatment groups 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.

To assess the heterogeneity of treatment effects across the subgroup levels, two Cox regression model will be fitted with the 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 x subgroup-variable

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

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

Power considerations for the interaction test of treatment and PD-L1 status

Calculations for different scenarios of PD-L1+ prevalence rates and assumed treatment effects in PD-L1 negative indicate that the power of the interaction test varies substantially and can be relatively low. For example with 337 events in the PD-L1+ FAS, a HR of 0.7 in the PD-L1+, and 80% PD-L1+ prevalence rate an interaction test will have power of 10% and 31%, when the HR in the PD-L1- group is 0.825 and 1.0, respectively. If the PD-L1+ prevalence is as low as 50%, power would be at 19% and 64%, when the HR in the PD-L1- group is 0.825 and 1.0, respectively.

16.1.4   Time of Follow-Up for OS

Analysis sets: PD-L1+ FAS / FAS

A Kaplan-Meier plot will be created for both treatment groups to assess the duration of follow-up for OS using the following censoring rules (reversing the OS censoring and event indicators):

Table 6   Censoring Rules for Duration of Overall Survival Follow-up

	Date of event / censoring	**Censoring**
Subjects alive or lost to follow-up	Time from randomization to last date known to be alive	No
Subjects who died	Time from randomization to date of death	Yes

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

16.2   Secondary Endpoint Analyses

Secondary efficacy analyses will be performed on the FAS and the PD-L1+ FAS. Secondary analysis of subject-reported outcomes / quality of life will be performed on the FAS.

To control the overall alpha level, the hierarchical testing strategy as described in Section 16 will be applied to test the secondary efficacy endpoints. With this testing strategy, if the one-sided p-value from the stratified log-rank test of the primary endpoint OS is below the significance level at the interim analysis according to O'Brien-Fleming boundaries based on a Lan-DeMets alpha spending function (see Table 1, Section 8.2) at the interim and the final analysis, respectively, the secondary efficacy endpoints will be tested with the specified order within the hierarchical testing procedure at the same nominal significance level as the primary endpoint if, and only if, the preceding null hypothesis within the specified hierarchical order was rejected. With this approach, the overall alpha level is controlled at 0.025 (one-sided). Using a hierarchical design to test secondary hypotheses, we will hold family-wise type I error rate of 2.5%, if we hold type I error rate at 2.5% for each of the secondary hypotheses within a pre-specified group-sequential design. Tamhane et al. (7) have summarized research on this topic in the introduction of a paper in 2012: *"Tamhane et al. (8) studied a two-stage group sequential procedure (GSP) for the problem of testing primary and secondary endpoints with the former acting as a gatekeeper for the latter. Hung et al. (9) were the first to study this problem. They compared three different strategies via*

simulation in terms of the type I error rate control for different values of ρ [=correlation coefficient between primary and secondary endpoint]. They showed that the strategy that tests the secondary null hypothesis at level α upon rejecting the primary null hypothesis does not control the error rate, but the strategy that tests the secondary null hypothesis at level α /2 is conservative. Finally, the strategy that tests both the primary and the secondary null hypotheses by using the same α - level critical boundary controls the error rate more accurately. Glimm et al. (10) independently obtained the same results as in Hung (9)". This shows that a group sequential procedure for a primary and a secondary endpoint controls the type I error rate, if we pre-specify a separate group sequential design for each of the primary endpoints.

To account for the two-stage group sequential design as applied in this study for the secondary efficacy endpoints, the repeated CI method as described in Section 16.1.1 will be used to construct the two-sided CI for the secondary efficacy endpoints of BOR and PFS at the interim analysis and the final analysis. In addition, the unadjusted 95% CIs of BOR and PFS will also be calculated at the interim and the primary analysis in order to provide comparable results based on the subgroup analysis.

A separate Imaging Data Management Plan and Data Transfer Plan will be created to summarize the details of the data structure and data delivery schedule of IERC assessment results.

16.2.1   Best Overall Response

Analysis sets: FAS / PD-L1+ FAS

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 IERC. 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 BOR. Clinical deterioration will not be considered as documented disease progression.

The following requirement is taken into account for confirmation of response:

- 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 summarizes the derivation rules described by Eisenhauer, et al. for the 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 BOR.

Table 7   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
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 as provided by the IERC in the analysis population. Subjects with BOR of non-CR/non- PD 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 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 alpha value as specified in Table 1, Section 8.2 will be applied to calculate the two-sided CI for the interim analysis and the final analysis, respectively.

The association of treatment and BOR will be tested by the General Association Statistic of the Cochran-Mantel-Haenszel test (CMH) (4) 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 and tumor response at least in one randomization stratum. The one-sided CMH test will be performed based on the significance test

level as specified for some example situations in Table 1, Section 8.2 at 75% interim analysis and the final analysis.

The stratified odds ratio in terms of OR will also be estimated along with its 95% CI to compare the treatment effect. The odds ratio is defined as the odds of OR with avelumab divided by the odds of OR with docetaxel. The Breslow-Day test will be used to check 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 (5).

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 as 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 (6).

In addition, the number and percentage of subjects with BOR of CR, PR, SD, PD, non-CR/non-PD, and Non-evaluable (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 (for IERC assessments: 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 anti-cancer therapy started before first post-baseline assessment

- SD of insufficient duration (<6 weeks after date of 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 too early".

The analysis based on IERC data will primarily consider radiological assessments because planned oncology reviews could not be conducted in a blinded manner. Hence sensitivity analysis are planned to integrate oncology reviews.

16.2.2   Progression Free Survival

Analysis sets: FAS / PD-L1+ FAS

The PFS time is defined as the time from date of randomization until date of the first documentation of PD or death by any cause (whichever occurs first). The tumor response will be

determined according to RECIST 1.1 and adjudicated by an IERC. Details on determination of the first disease progression date are specified in the IERC charter.

PFS time (in months) = (Date of PD or death - date of randomization + 1)/ 30.4375 (months)

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 anti-cancer treatment prior to an event, or for subjects with an event after two or more consecutive 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.

Last adequate tumor assessment is defined as the last tumor assessment result that is not "NE" or "NA".

The censoring and event date options to be considered for the PFS and Duration of Response (DR) analysis are presented in Table 8.

The analysis of PFS will be analogous to that for OS time as described in section 16.1.1. The alpha value as specified in Table 1, Section 8.2 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 PFS rate at 6, 9, 12 and 15 months will be estimated with corresponding two-sided 95% CI.

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:

- Ongoing in the study without an event
- 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 page prior to the analysis cut-off;

- Subjects with the last contact date > 14 weeks prior to the analysis cut-off date (duration of 14 weeks is based on the assessment schedule of every 3 months for survival follow-up interval + 1 week window).

Table 8   **Outcome and Event Dates for PFS and DR 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 date of randomization	Date of PD or death	Event
PD or death after 2 or more missing or inadequate post-baseline tumor assessments	Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
No PD and no death	Date of last adequate tumor assessment[b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored
Treatment discontinuation due to 'Disease progression' without documented progression	Not applicable	Information is ignored. Outcome is derived based on documented progression only.
New anti-cancer therapy given	Date of last adequate tumor assessment [b] documenting no PD before new anti-cancer therapy is given or missed tumor assessments	Censored

[a] However if the patient dies ≤12 weeks after date of randomization the death is an event with date on 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.

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

The analysis based on IERC data will primarily consider radiological assessments because planned oncology reviews could not be conducted in a blinded manner. Hence sensitivity analysis are planned to integrate oncology reviews.

16.2.3   Health-Related Quality of Life

Analysis sets: HRQoL Analysis Set / PD-L1+ HRQoL Analysis Set

Health-related quality of life (HRQoL) will be assessed by the EuroQOL 5-dimensions questionnaire (EQ-5D), the European Organization for Research and Treatment of Cancer

(EORTC) QLQ-C30 supplemented by the lung cancer module QLQ-LC13 questionnaire. The analyses specified below will be performed for HRQoL analysis set and PD-L1+ HRQoL analysis set. Further exploratory and sensitivity analyses on HRQoL assessment will be detailed in an addendum to this analysis plan which will include but not limited to assessing time assessment bias (the two treatment arms have different schedule of assessment).

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). It also includes a visual analogue scale ranging from 0 to 100 for self-rated health status. A higher score indicates 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 incorporates five functional scales (physical, role, cognitive, emotional, and social), three symptom scales (fatigue, pain, and nausea and vomiting), a global health status / QoL scale, and a number of single items assessing additional symptoms commonly reported by cancer subjects (e.g. dyspnea, loss of appetite, insomnia, constipation, and diarrhea), and financial impact of the disease.

All of the scales and single-item measures range in score from 0 to 100. A high scale score represents a higher response level. Therefore, a high score for a functional scale represents a high/healthy level of functioning, a high score for the global health status / QoL represents a high QoL, but a high score for a symptom scale/item represents a high level of symptomatology / problems. The scoring procedure for each of the scales is the same and consists of computing the raw score (RS) and then computing the actual scale score (S) by making a linear transformation to standardize the score to values from 0 to 100 as shown below.

$$\text{Raw Score} = RS = (I_1 + I_2 + \ldots + I_n) / n$$

For **Functional scales**:

$$\text{Score} = 100 \times [1 - (RS - 1) / \text{Range}]$$

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

$$\text{Score} = 100 \times [(RS - 1) / \text{Range})]$$

Where I_1, I_2, … I_n are the individual items and Range is the difference between the maximum possible value of RS and the minimum possible value. The range of RS equals the range of the item values. Most items are scored 1 to 4, giving Range = 3. The exceptions are the items contributing to the Global Health Status / QoL, which are 7-point questions with Range = 6.

Table 9   QLQ-C30 Score System

Scale	Item range*	Item numbers
Global Health Status / QoL		
Global Health Status / QoL	6	29,30
Functional Scales		
Physical Functioning	3	1,2,3,4,5
Role Functioning	3	6,7
Emotional Functioning	3	21,22,23,24
Cognitive Functioning	3	20,25
Social Functioning	3	26,27
Symptom Scales / Items		
Fatigue	3	10, 12, 18
Nausea and vomiting	3	14, 15
Pain	3	9, 19
Dyspnoea	3	8
Insomnia	3	11
Appetite loss	3	13
Constipation	3	16
Diarrhoea	3	17
Financial difficulties	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.

The QLQ-LC13 questionnaire comprises 13 lung cancer-specific questions incorporated into one multi-item scale designed to evaluate lung cancer related symptoms [cough and hemoptysis (one item each), dyspnea (three items)], treatment related side-effects [sore mouth or tongue, dysphagia, hair loss, tingling hands, and feet (one item each)], pain (three items), and pain medication (one item). All items are rated on a 4-point scale, with 1 representing no symptom/problem at all, and 4 representing worst symptom/problem.

Table 10   QLQ-LC13 Score System

Symptom Scales / Items	Item numbers
Dyspnoea[†]	3, 4, 5
Coughing	1
Haemoptysis	2
Sore mouth	6
Dysphagia	7
Peripheral neuropathy	8
Alopecia	9
Pain in chest	10

Pain in arm or shoulder	11
Pain in other parts	12
Take any medication for pain	13

For QLQ-C30 questionnaire, 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.

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

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

- HRQoL questionnaires compliance and full completion rates

  Compliance and full completion rates will be summarized for EQ-5D, QLQ-C30 Global Health status and functional scales and QLQ-LC13 questionnaire for each scheduled visit, considering the following definitions.

  $$\%Compliance = 100 \times \frac{\text{number of subjects with at least one HRQoL questionnaire available}}{\text{number of subjects for whom a HRQoL questionnaire is expected}}$$

  $$\%Full\ completion\ rate = 100 \times \frac{\text{number of subjects with all items in HRQoL questionniare available}}{\text{number of subjects for whom HRQoL questionnaire is expected}}$$

  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. Additionally, change from baseline will also be reported at scheduled visit by 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 using a line plot with time point (scheduled visit) on the x-axis, mean change score on the y-axis, and standard errors around each mean score. Separate lines will be presented for each treatment arm.

  If there are multiple complete assessments for any scheduled visit during the treatment phase, 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 first assessment 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 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 will be analyzed as a continuous variable. The overall significance of the difference between the MMRM trajectories for the treatment arms will be tested at a significance level of 0.05 (two-sided). An unstructured covariance structure will be used, if possible, though a simpler covariance structure will be used if the model does not converge.

- Subject level change from baseline

  For the HRQoL domain scores analyzed using MMRM, an empirical cumulative distribution function (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 function 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 domain score change equivalent to the Minimum Important Difference (MID) and such change is confirmed by no further improvement of domain score or no further available HRQoL data due to death.

  The MID is defined as:

  – <u>QLQ-C30 physical function score</u>: A change of 10 normalized score points;

  – <u>QLQ-LC13 symptom scale</u>: An change of 10 normalized points;

  The deterioration of each domain score is defined as follows:

  – <u>QLQ-C30 physical function scale score</u>: change from baseline < –MID;

  – <u>QLQ-LC13 symptom scale</u>: change from baseline > MID;

The deterioration of each HRQoL score needs to be confirmed with the following criteria for the occurrence to be *definitive* deterioration:

- <u>QLQ-C30 physical function scale score</u>: After the HRQoL score deterioration, no increase >= MID compared to the lowest score or no available HRQoL data due to death;

- <u>QLQ-LC13 symptom scale</u>: After the HRQoL score deterioration, no decrease >= MID compared to the highest score or no available HRQoL data due to death;

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

For subjects without any DD of the respective HRQoL assessment, or with a death date >= 12 weeks after the last HRQoL assessment or with two or more missing post-baseline HRQoL assessment, TUDD will be censored on the date of last corresponding HRQoL assessment or date of randomization whichever is later. For subjects who start new anti-cancer treatment prior to a DD, TUDD will be censored on the date of last HRQoL assessment before anti-cancer therapy is given. The time window of 12 weeks corresponds to the time window used for the PFS endpoint based on the 6-week HRQoL assessment schedule.

Only DDs up to 5 days after treatment discontinuation will be considered. Date of treatment discontinuation is collected on the "Treatment Termination" eCRF page.

Missing Values and Imputation Rules

For a multi-item subscale on the QLQ-C30 and QLQ-LC13 score system, if more than 50% of the questions that comprise this subscale have missing values, this subscale will be considered missing.

Missing data will be retained as observed, and no imputation of HRQoL scores will be conducted.

16.2.4   Sensitivity Analyses of Secondary Efficacy Endpoints

Analysis sets: FAS / PD-L1+ FAS

For the secondary efficacy endpoint of PFS and 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 PFS and BOR.

- PFS and 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

CCI

16.2.5   Time of Follow-up for PFS

Analysis sets: FAS

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

Table 11   Reverse Censoring Rules for Duration of PFS Follow-up

	Date of event / censoring	**Censoring**
Subjects alive or lost to follow-up	Time from randomization to last date known to be alive	No
Subjects who died or had PD	Time from randomization to date of first PD or death	Yes

Kaplan-Meier estimates will be presented by treatment group together with the median time of follow-up for PFS. In particular, the follow-up rate at 3, 6, 9 and 12 months will be estimated with corresponding two-sided 95% CIs.

16.2.6   Subgroup Analysis of Secondary Efficacy Endpoints

Analysis sets: FAS / PD-L1+ FAS

Analysis on subgroups as defined in Section 10 will be performed for the secondary efficacy endpoints, BOR and PFS assessments.

All the subgroup analyses will be exploratory; no adjustment for multiplicity will be performed. In the case of a low number of subjects within a category (< 25 subjects, which is about 5% of the randomized PD-L1+ population), the categories may be pooled when meaningful.

All the subgroup analyses will be performed as unstratified. The unadjusted 95% CI will be calculated for all subgroup analyses of BOR, PFS and subject-reported outcomes/ quality of life assessments.

The subgroup analysis of PFS will be analogous to that of subgroup analysis of OS time as described in section 16.1.1.

For the subgroup analysis of BOR, the association of treatment and BOR will be tested using the two-sided CMH test per subgroup level. The ORR along with the two-sided exact Clopper-Pearson 95% CIs will be calculated for each subgroup.

In addition, 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.

16.3   Other Endpoint Analyses

16.3.1   CCI

CCI

██████████████████████████████████ CCI █████████████████████████████

██
██
████████████████████████████████████

██
███

██
█████████████████████████

16.3.2 CCI ████████████████████████

CCI ██████████████████████████████

██
██
██

▪ ██
  ██████████████████████████

██
██
██████████████████████████

▪ ██
  ██████████████████████████████████

██
██
██
██
██
██

CCI ███████████   ██

CCI ██████████████████████████████

██
█████████████████████████████████

██
██
██
███

CCI

CCI

CCI	

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

17   Safety Evaluation

Analysis sets: Safety Analysis Set / PD-L1+ Safety 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 and the PD-L1+ Safety analysis set and according to the as-treated principle. The safety endpoints will be tabulated using descriptive statistics.

17.1   Adverse Events

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 11.

All analyses described in Section 17.1 and 17.2 will be based on TEAEs (started during the on-treatment period) if not otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not). A separate listing including AEs started after the on-treatment period will also be provided.

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with trial treatment = Related) reported by the investigator and those of missing or unknown relationship (i.e. no answer to the question "Relationship with trial 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):** immune related adverse events according to case definition classified by medical review. Details are included in Table 16 in Appendix V.

- **Infusion Related Reactions (IRR):** IRRs are identified based on a list of MedDRA PTs. The detailed criteria of the timing relationship to infusion are specified in Table 17 in Appendix V.

Unless otherwise stated adverse events will be displayed in terms of frequency tables by treatment group: PT and primary SOC by decreasing frequency based on the avelumab treatment arm.

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

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.

17.1.1　　　　All Adverse Events

Adverse events will be summarized by worst grade according to National Cancer Institute – Common Terminology Criteria for Adverse Events (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 tables will be created:

- The overall summary of AEs table will include the following per 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 the 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 All TEAEs / Related TEAEs / Grade ≥3 TEAEs / Related Grade ≥3 TEAEs

- Non-serious TEAEs with frequency ≥ 5% in any treatment arm by SOC and PT

17.1.2 Adverse Events Leading to 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 discontinuation of avelumab / docetaxel by SOC and PT

- Related TEAEs leading to discontinuation of avelumab / docetaxel by SOC and PT

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

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

17.2.1 Deaths

All deaths, deaths within 30 days after last dose of study drug, deaths 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 dose of study treatment

- Flag for death within 60 days of first dose of study treatment

17.2.2 Serious Adverse Events

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

- Incidence of serious AEs by SOC and PT

- Incidence of avelumab-related serious AEs by SOC and PT

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

17.2.3 Other Significant Adverse Events

In order to thoroughly and consistently analyze treatment emergent irAEs, a two-level approach is performed as follows. The details are specified in Appendix V.

> **Level 1:** A MedDRA Preferred Term (PT) query is performed for each event category (i.e., immune-mediated rash, colitis, pneumonitis, hepatitis, nephritis and renal dysfunction, endocrinopathies and other immune-mediated adverse reactions). These irAEs are referred as "potential" irAEs and will be summarized for both treatment arms.

> **Level 2:** AEs identified by the MedDRA PT queries will then be medically reviewed using pre-defined case definitions for immune-mediated adverse reactions. These irAEs will be summarized for the avelumab arm only.

The following overall frequency tables will be prepared for treatment emergent irAEs for both potential irAEs and irAEs per sponsor medical review. The cluster is a compilation of PTs that are categorized by immune-related event of special interest as specified in Appendix V.

- The overall summary of 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 $\geq$ 3, by Cluster and PT

- irAEs leading to permanent treatment discontinuation by Cluster and PT

- irAEs by Cluster and PT and worst grade

The listing of all potential irAEs will be provided with the relevant information with a flag for irAEs with onset outside of the on-treatment period. The irAEs that are identified per sponsor medical assessment and are only applicable to avelumab treatment arm will also be flagged in the listing. A separate listing of irAEs with onset after the on-treatment period will also be provided.

The following overall frequency tables will be prepared for treatment emergent IRRs, by treatment group:

- The overall summary of IRR table will include the following categories:
  - All IRRs
  - Related IRRs
  - Serious IRRs
  - Related serious IRRs
  - IRRs, Grade $\geq$ 3
  - Related IRRs, Grade $\geq$ 3
  - IRRs leading to permanent treatment discontinuation
  - Related IRRs leading to permanent treatment discontinuation
  - IRRs leading to death
  - Related IRRs leading to death
- IRRs leading to death, by PT

- Related IRRs leading to death, by PT

- IRRs, by PT

- IRRs, Grade $\geq$ 3, by PT

- Related IRRs, by PT

- Related IRRs, Grade $\geq$ 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

- Time related to first onset of an IRR (infusion 1, infusion 2, infusion 3, infusion 4 or later)

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

17.2.4   CCI

CCI

17.3   Clinical Laboratory Evaluation

17.3.1   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).

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

Quantitative data will be examined for trends using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each visit over time. End of Treatment visit will be summarized separately. The changes computed will be the differences from baseline. Qualitative data based on reference ranges will be described according to the categories (i.e. Low, Normal, High).

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

Unscheduled laboratory 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 the worst grade for NCI-CTC gradable parameters and normality shift table for NCI-CTC non-gradable parameters during the on-treatment period.

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

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

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

- Lymphocyte count decreased:
    - derived absolute count does not meet Grade 2-4 criteria, and
    - % value < % LLN value, and
    - derived absolute count ≥ 800/mm3
- Neutrophil count decreased
    - derived absolute count does not meet Grade 2-4 criteria, and
    - % value < % LLN value, and
    - derived absolute count ≥ 1500/mm3

For **calcium**, CTCAE grading is based on Corrected Calcium. 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 tests 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, ALT ≥ 5×ULN, ALT ≥ 10×ULN, ALT ≥ 20×ULN
- AST ≥ 3×ULN, AST ≥ 5xULN, AST ≥ 10×ULN, AST ≥ 20×ULN
- (ALT or AST) ≥ 3×ULN, (ALT or AST) ≥ 5×ULN, (ALT or AST) ≥ 10×ULN, (ALT or AST) ≥ 20×ULN
- TBILI ≥ 2×ULN
- Concurrent ALT ≥ 3×ULN and TBILI ≥ 2×ULN

- Concurrent AST $\geq$ 3×ULN and TBILI $\geq$ 2×ULN

- Concurrent (ALT or AST) $\geq$ 3×ULN and TBILI $\geq$ 2×ULN

- Concurrent (ALT or AST) $\geq$ 3×ULN and TBILI $\geq$ 2×ULN and ALP > 2×ULN

- Concurrent (ALT or AST) $\geq$ 3×ULN and TBILI $\geq$ 2×ULN and (ALP $\leq$ 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 $\geq$10×ULN will also appear in the categories $\geq$5×ULN and $\geq$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 subjects with a post-baseline TBILI $\geq$ 2×ULN, ALT$\geq$ 3×ULN or AST $\geq$ 3×ULN will be provided.

Parameters with NCI-CTC 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 CTCAE grade >=0, grade >=3 and grade>=4 laboratory abnormalities during the on-treatment period.

- The shift table will summarize baseline CTCAE grade versus the worst on-treatment CTCAE grade. The highest CTCAE grade during the on-treatment period is considered as the worst grade for the summary.

The above analyses apply to hematology and chemistry evaluations which can be graded per CTCAE, 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 (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-CTC grades not available:

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

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

- Number 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, Red Blood Cell (RBC), Reticulocytes, Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin Concentration (MCHC).

- Serum Chemistry:

  Chloride, C-Reactive Protein, Lactate Dehydrogenase (LDH), Total Protein, Total Urea, Uric Acid.

17.3.2   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 into 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.

17.4   Vital Signs

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

Table 12   Categories of Change from Baseline for Vital Sign Parameters

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

bpm = beats per minute for heart rate and breaths per minute for respiration rate; DBP=diastolic blood pressure; SBP=systolic blood pressure.

All vital sign parameters will be examined for trends using descriptive statistics (mean, StDev, 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.

For each subject the worst on-treatment value will be calculated. Missing values will define a separate category. A summary of maximum shift from baseline by category will be provided by treatment arm. A listing of maximum change from baseline per subject will also be provided.

17.5   Other Safety or Tolerability Evaluations

17.5.1   ECG

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 the worst potentially clinically significant abnormalities (PCSA) for ECG parameters will be summarized during the on-treatment period. Each subject will be counted only once within each category. As ECG assessments are only performed during screening and at the Discontinuation/End-of-Treatment visit, the denominator to calculate percentages for each PCSA category is the number of subjects with Discontinuation/End of Treatment visit. The PCSA criteria are provided in Table 13 below.

Table 13   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	Interval >450 ms and interval ≤ 480 ms Interval >480 ms and interval ≤ 500 ms Interval >500 ms
QTcF and QTcB change from baseline	Increase from baseline > 30 ms and ≤ 60 ms Increase from baseline > 60 ms

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

A listing of abnormal 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.

17.5.2   ECOG Performance Status

The ECOG shift from baseline to highest score during the on-treatment period will be summarized. 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.

17.5.3 CCI ████████████

CCI ████████████████████████

██
██
██
██
██

CCI ██████   ████████████████████████████████

CCI ██████████████   ██████████   ███   ██████████

████████   ████   ███   ████████
███   ████   ███   ███
████████   ████████   ███   ████████
████████   ████   ███   ███
████████   ████████   ████████   ████████
████████   ████████   ███   ████████████

██
██

CCI ██████   ████████████████████████████████

CCI ██████   ████████   ████████████████████████████████████

███   ████████████████████████   ████████████████████████████
████████   ████████████████████
███   ████████████████████████   ████████████████████████████
████████   ████████████

███   ████████████████████████████   ████████████████████████████
████████   ████████
████████   ████████████████████████   ████████████████████████████
████████   ████████████████████████   ████████████████████
           ████████████████

████████   ████████████████████████   ████████████████████████████
████████   ████████████████████████   ████████████████████████████
                                        ████████████████████████

████████   ████████████████████████   ████████████████████████████
████████   ████████████████████████   ████████████████████████████
           ████████████████████████   ████████████████████████
           ████████████

████████   ██ ████████ ████████ ██████ ████ ████   ████████████████████████████
████████   ████████████████████████   ████████████████████████████
           ████████████████████████   ████████████████████████

████████████████████████████
█ ██

CCI
███
███
███
███
█████████████

CCI █████████   ██

 CCI ████████████████████   ███████████   ███████████████

██████████   ██████   ███████

████   ████   ████

████   ████   ████████

██████████   ████   ████

████████   ████████   ████████

████████   ████████   ████████
██

CCI █████████   ███████████████████████████████████████

 CCI ████████   ████████████   ██

██████   ██████████████████████████████████   ████████████████████████████████████
████████   ████████████████████████████

██████   ████████████████████████████████████   ████████████████████████████████████
████████   ████████████████████████████

██████   ████████████████████████████████████   ████████████████████████████████████
████████   █████████   ████

██████████ ████████████████████████████████████   ████████████████████████████████████
██████████ ████████████████████████████████████   ████████████████████████████████████
            ████████   ████████████████████████████████████
                                                     ████

██████████ ████████████████████████████████████   ████████████████████████████████████
████████   ████████████████████████████████████   ████████████████████████████████████
            ████████████████████████████████████   ████████████████████████████████████
            ██████████████████████████   ████

██████████ ██ █ ████████ ████████ ██████ ██████   ████████████████████████████████████
████████   ████████████████████████████████████   ████████████████████████████████████
            ████████████████████████████████████   ████████████████████████████████████
            ██████████   ████
██

CCI ███
███
██████████████████

█████████████████████████

CCI

CCI

18   References

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

2. Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics 1982;38:29-41.

3. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons 1980.

4. Cochran WG. Some Methods for Strengthening the Common Chi-square Tests. Biometrics 1954;10:417-51.

5. Agresti A. A Survey of Exact Inference for Contingency Tables. Statistical Science 1992;7:131–77.

6. Agresti A. An Introduction to Categorical Data Analysis (2 ed.). New York: John Wiley & Sons, 2007.

7. Tamhane AC, Wu Y, Mehty CR. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. Statistics Med 2012;31:2027-40.

8. Tamhane AC, Mehta CR, Liu L. Testing a Primary and a Secondary Endpoint in a Group Seuqential Design. Biometrics 2010;66:1174-84.

9. Hung J, Wang SJ, O'Neill R. Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials. J Biopharm Statistics 2007;17:1201-10.

10. Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Statistics Med 2009;29:219-28.

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

12. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. DOI 10.1200/jco.1998.16.1.139 J Clin Oncol. 1998 Jan;16(1):139-44.

13. AS Pickard, MP Neary, D Cella : Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer Health Qual Life Outcomes 5: 70,2007

14. Maringwa JT, Quinten C, King M, Ringash J, Osoba D, Coens C, et al. Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials.Support Care Cancer2011;19:1753-60

19 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   Important and Clinically Important Protocol Deviations by Programming Check and Medical Review

Description of Protocol Deviation	Deviation Code	Clinically Important PD?	Proposed check / comment
Inclusion/Exclusion Criteria			
Subject did not meet inclusion criterion #3: i.e. Availability of a formalin-fixed, paraffin-embedded block containing tumor tissue or unstained tumor slides suitable for PD-L1 expression assessment	INCEXC01		Programming check and Medical review required
Subject did not meet inclusion criterion #4; i.e. Tumor determined not be evaluable for PD-L1 expression per the evaluation of the central laboratory	INCEXC02	Yes	Programming check and Medical review required
Subject did not meet inclusion criterion #5; i.e. Subjects with histologically confirmed Stage IIIb/IV or recurrent NSCLC who have experienced disease progression	INCEXC03	Yes	Programming check and Medical review required
Subject did not meet inclusion criterion #6; i.e. Subjects must have progressed during or after a minimum of 2 cycles of 1 course of a platinum-based combination therapy administered for the treatment of metastatic disease (#6a) or Subjects must have progressed within 6 months of completion of a platinum-based adjuvant, neoadjuvant, or definitive chemotherapy, or concomitant chemoradiation regimen for locally advanced disease (#6b)	INCEXC04	Yes	Programming check and Medical review required
Subject did not meet inclusion criterion #11 i.e. Subject does not have adequate hepatic function.	INCEXC05		Programming check and Medical review required
Subject met exclusion criterion #2, i.e. Systemic anticancer therapy administered after disease progression during or following a platinum-based combination	INCEXC06	Yes	Programming check and Medical review required
Subject met exclusion criterion #3, i.e. Subjects with non-squamous cell NSCLC whose disease harbors EGFR mutation(s) and / or anaplastic lymphoma kinase (ALK) rearrangement will not be eligible for this trial. Subjects of unknown ALK and / or EGFR mutation status will require testing at screening.	INCEXC07	Yes	Programming check and Medical review required
Subject met exclusion criterion #4; i.e. Subject has received prior therapy with any antibody / drug targeting T-cell coregulatory proteins (immune checkpoints) such as PD-1, PD-L1, or cytotoxic T-lymphocyte antigen-4 (CTLA-4)	INCEXC08	Yes	Programming check and Medical review required.
Subject met exclusion criterion #5; i.e. Subject was found to be receiving 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	Yes	Programming check and Medical review required
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		Programming check and Medical review required

Avelumab
EMR 100070-004

Avelumab in Non-Small Cell Lung Cancer

Description of Protocol Deviation	Deviation Code	Clinically Important PD?	Proposed check / comment
Inclusion/Exclusion Criteria			
Subject met exclusion criterion #7: Subject has received immunosuppressive agents and were not tapered off these drugs before initiation of the study treatment (with the exception of subjects with adrenal insufficiency, who may continue corticosteroids at physiologic replacement dose, equivalent to < 10 mg prednisone daily)	INCEXC11		Programming check and Medical review required
Subject met exclusion criterion #8 i.e. presence of brain metastases, except those meeting the following criteria: a. Brain metastases have been treated locally, and b. No ongoing neurological symptoms that are related to the brain localization of the disease (sequelae that are a consequence of the treatment of the brain metastases are acceptable)	INCEXC12		Programming check and Medical review required
Subject met exclusion criterion #12: Subject has an active autoimmune disease that might deteriorate when receiving an immune-stimulatory agent	INCEXC14		Programming check and Medical review required
Subject met exclusion criterion #14: Subject has known severe hypersensitivity reactions to monoclonal antibodies (Grade ≥ 3 NCI-CTCAE v 4.03), any history of anaphylaxis, or uncontrolled asthma (that is, 3 or more features of partially controlled asthma)	INCEXC15		Programming check and Medical review required
Subject met exclusion criterion #19: pregnant or lactation	INCEXC16		Programming check and Medical review required
Any other inclusion criteria response "No" and/or any "Yes" response to any other exclusion criteria and subject was subsequently randomized	INCEXC17		Programming check and Medical review required
Informed Consent/ Subject information			
Subject did not sign ICF and was enrolled.	INFCON01		Programming check and medical review required
Investigational Product			
Avelumab premedication was not administered prior to IP dose.	INVPRO09		Programming check and Medical review required
Randomization			
Patient randomized more than once	RANDOM02		Programming check and medical review required

82/86

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

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

Fatigue

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

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 Review for Assessment of Immune-Related AEs and Definition of Infusion Related Reactions

In order to thoroughly and consistently analyze potential immune-mediated adverse events (AEs), a two-level approach is proposed including:

1. A MedDRA Preferred Term (PT) query is proposed for each event category (i.e., immune-mediated rash, colitis, pneumonitis, hepatitis, nephritis and renal dysfunction, endocrinopathies and other immune-mediated adverse reactions).

2. AEs identified by the MedDRA PT queries will then be medically reviewed using pre-defined case definitions for immune-mediated adverse reactions.

Level 1:

To identify potentially immune-mediated AEs, the MedDRA PT queries will be used to search for AEs of interest in the clinical database. The proposed event categories such as:

Immune-mediated rash, Immune-mediated colitis, Immune-mediated pneumonitis, Immune-mediated hepatitis, Immune-mediated nephritis and renal dysfunction, Immune-mediated endocrinopathies (Thyroid disorders: Hypothyroidism, Hyperthyroidism, and Thyroiditis), Immune-mediated endocrinopathies (Adrenal insufficiency, Immune-mediated endocrinopathies (Type 1 Diabetes Mellitus), Immune-mediated endocrinopathies (Pituitary dysfunction), Immune-mediated endocrinopathies (Hypogonadism), Other immune-mediated adverse events. Further details e.g. MeDDRA PT queries are regularly updated based on the current MeDRA version.

In order to standardize the MedDRA PT queries as much as possible, High Level Terms (HLT) and Standardized MedDRA Queries (SMQ) were used whenever a choice, that was considered reflective of the events of interest, was available.

Level 2:

In a second level (medical review), the potential immune-mediated AEs identified from the search performed at Level 1, will be reviewed by qualified medical personnel to determine whether the AE meets the criteria (case definition) for an immune-mediated adverse reaction based on the following algorithm:

Table 18 Algorithm for immune-related adverse reactions

Criteria	Description
Onset	AE onset after 1st avelumab administration until up to 90 days after last dose
Duration	AE does not spontaneously resolve (i.e., without corticosteroids/ immunosuppressant treatment) within 7 days after onset
Immunosuppressive therapy	AE treated with corticosteroid or other immunosuppressant therapy. *For endocrinopathies only*: AE required hormone replacement* and /or (corticosteroid or other immunosuppressive therapy)

Etiology	No other clear etiology or Histopathology/biopsy consistent with immune-mediated event
All criteria listed in the left column need to be fulfilled for an event to meet the case definition of immune-mediated reaction.	
*Hormone replacement will be evaluated for specific endocrinopathy disorders only as follows: • Thyroid disorders (HLT): Thyroid therapy (ATC codes (H03A, H03B)) • Diabetes mellitus (including hyperglycaemia): Insulin (ATC code A10A)	

Infusion related reactions are identified based on a list of MedDRA PTs and criteria on the timely relationship according to Table 19.

Table 19   Criteria for infusion related reactions

Infusion related reactions	**Reactions - Considered when onset is on the day of avelumab infusion (during or after the infusion) or the day after the avelumab infusion (irrespective of resolution date):** • Infusion related reaction • Drug hypersensitivity • Anaphylactic reaction • Hypersensitivity • Type 1 hypersensitivity **Signs and Symptoms - occurring on the day of avelumab infusion (during or after the infusion) and resolved with end date within 2 days after onset** • Pyrexia • Chills • Flushing • Hypotension • Dyspnea • Wheezing • Back pain • Abdominal pain • Urticaria