STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-A317-303
Study Protocol Title:	A Phase 3, Open-Label, Multicenter, Randomized Study to Investigate the Efficacy and Safety of BGB-A317 (Anti-PD1 Antibody) Compared with Docetaxel in Patients with Non-Small Cell Lung Cancer Who Have Progressed on a Prior Platinum-Containing Regimen
Date:	October 16, 2020
Version:	1.0
NCT ID:	NCT03358875

SIGNATURE PAGE

Author: [redacted]
Senior Principal Statistician,
Statistics and Data Science

[redacted]

Approval

[redacted]
Vice President,
Statistics and Data Science

[redacted]

[redacted]
Executive Director,
Clinical Development I/O

[redacted]

[redacted]
Medical Director,
Clinical Development

[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	Antidrug antibody
AE	adverse event
BOR	best overall response
BP	blood pressure
CAP	China Asia Pacific
CI	confidence interval
CR	complete response
CRF	case report form
CT	computerized tomography
CTCAE	Common Terminology Criteria for Adverse Events
C_{trough}	minimum observed plasma concentration
DCR	disease control rate
DOR	duration of response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
GEP	gene expression profiling
ITT	Intent-to-treat
irAE	immune-related AE
KM	Kaplan-Meier
MedDRA®	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	objective response rate
OS	overall survival
PD	progressive disease
PD-1	programmed cell death-1
PFS	progression-free survival
PK	pharmacokinetics
PR	partial response
PT	preferred term
RECIST	Response Evaluation Criteria in Solid Tumors

ROW	rest of the world
SAE	serious adverse events
SAF	Safety Analysis Set
SAP	statistical analysis plan
SD	stable disease
SOC	system organ class
TC	tumor cell
TEAE	treatment-emergent adverse event
TLG	table listing and graph
TMB	tumor mutational burden

1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report results for BGB-A317-303: A Phase 3, Open-Label, Multicenter, Randomized Study to Investigate the Efficacy and Safety of BGB-A317 (Anti-PD1 Antibody) Compared with Docetaxel in Patients with Non-Small Cell Lung Cancer Who Have Progressed on a Prior Platinum-Containing Regimen. The focus of this SAP is for the planned primary, secondary and exploratory analysis specified in the study protocol.

The analysis details for exploratory biomarker analyses are not described within this SAP. Separate analysis plans may be completed for these analyses and may be attached to the clinical study report.

Reference materials for this statistical plan include the protocol amendment BGB-A317-303 (version 3.0, dated as 9 March 2020). If the protocol is amended or updated, then appropriate adjustments to the SAP may be made if they are related to the planned analyses.

The SAP described hereafter is a priori plan. This is an open label study with a planned interim analysis, and the SAP will be finalized and approved before interim analysis. Statistical programming may occur as study data accumulate in order to have analysis programs ready prior to database lock.

2 STUDY OVERVIEW

Study Design

This is a randomized, open-label multicenter Phase 3 study designed to compare the efficacy and safety of tislelizumab versus docetaxel as treatment in the second- or third-line setting in patients with NSCLC. The primary endpoint of the study is OS in both the ITT and PD-L1 positive analysis sets.

Patients must have histologically confirmed, locally advanced or metastatic NSCLC (squamous or non-squamous). Histology of NSCLC (squamous or non-squamous) will be confirmed at the investigator's site. Patients with known epidermal growth factor receptor (EGFR) mutation or anaplastic lymphoma kinase (ALK) rearrangement are ineligible for the study. Documentation of wild type EGFR by tissue-based test is required for non-squamous patients to enter the study. Archival tumor tissues (not restricted to pretreatment) will be collected for biomarker analysis at a central laboratory. If archived formalin-fixed paraffin-embedded (FFPE) tissue is not sufficient, a fresh biopsy sample will be mandatory.

Patients must have been treated with at least one platinum-containing regimen but not more than 2 prior lines of systemic chemotherapy and have disease progression during or following chemotherapy treatment. Patients who progressed or have disease recurrence during or after neoadjuvant or adjuvant therapy with platinum-containing regimen (counted as one line of therapy) within 6 months are eligible to enroll into the study.

After signing an informed consent form and screening for eligibility, patients will be randomized in a 2:1 ratio to receive tislelizumab or docetaxel in the following 2 arms:

Arm A: Tislelizumab 200 mg intravenously (IV) once every 3 weeks

Arm B: Docetaxel 75 mg/m^2 IV once every 3 weeks

Randomization will be stratified by histology (squamous versus nonsquamous), line of therapy (second versus third), and PD-L1 expression (≥25% TC versus <25% TC).

Administration of docetaxel and tislelizumab will continue until disease progression, as assessed by investigator per RECIST v1.1, unacceptable toxicity, or withdrawal of informed consent, whichever occurs first. Patients receiving tislelizumab will be permitted to continue tislelizumab treatment beyond progression by RECIST v1.1 if clinical benefit is seen in the absence of symptomatic deterioration and unacceptable toxicity per investigator's discretion.

Tumor response will be evaluated by investigators every 9 weeks (± 7 days) during Year 1 and every 12 weeks (± 7 days) from Year 2 onwards based on RECIST v1.1. If a patient discontinues study treatment due to reasons other than disease progression, then tumor assessments should continue to be performed as scheduled until the start of new anticancer therapy, disease progression by RECIST v1.1, death, loss to follow-up, withdrawn consent, or until the study closes, whichever occurs first. Any new anticancer therapy after discontinuation of study treatment from both arms will be documented.

To determine the PK properties of tislelizumab and host immunogenic response to tislelizumab, blood samples will be collected at various time points.

Patients will be evaluated for adverse events (AEs) and serious adverse events (SAEs) occurring up to 30 days after the last dose of either study drug or initiation of new anticancer therapy, whichever occurs first, (all severity grades according to NCI-CTCAE v4.03) and all immune-related AEs (irAEs) (tislelizumab arm only) occurring up to 90 days after the last dose of tislelizumab, regardless of whether or not the patient starts a new anticancer therapy. All drug related SAEs will be recorded by the investigator after treatment discontinuation until patient death, withdrawal of consent, or loss to follow up, whichever occurs first.

3 STUDY OBJECTIVES

3.1 PRIMARY OBJECTIVES

- To compare the efficacy, as measured by overall survival (OS), of tislelizumab with docetaxel in the second- or third-line setting in patients with non-small cell lung cancer (NSCLC) who have progressed on a prior platinum-containing regimen. A comparison of the treatment arms will be performed in:
    - The intent-to-treat (ITT) analysis set

- The programmed cell death protein ligand-1 (PD-L1) positive analysis set, where PD-L1 positive is defined as ≥25% of tumor cells (TCs) with PD-L1 membrane staining via the Ventana SP263 assay

3.2 SECONDARY OBJECTIVES

- To compare the efficacy of tislelizumab and docetaxel as measured by objective response rate (ORR), duration of response (DoR), and progression-free survival (PFS) per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 in:
  - The ITT analysis set
  - The PD-L1 positive analysis set
- To compare health-related quality of life (HRQoL) between tislelizumab and docetaxel arms
- To evaluate the safety and tolerability of tislelizumab versus docetaxel

3.3 EXPLORATORY OBJECTIVES

- To compare tumor assessment outcomes (ie, disease control rate [DCR] and clinical benefit rate [CBR]) between tislelizumab and docetaxel assessed by investigator per RECIST v1.1
- To explore potential predictive biomarkers for efficacy including but not limited to PD-L1 expression, tumor mutational burden (TMB), gene expression profile (GEP), and tumor-infiltrating immune cells
- To characterize the pharmacokinetics (PK) of tislelizumab in patients with NSCLC
- To determine host immunogenicity to tislelizumab in patients with NSCLC

4 STUDY ENDPOINTS

4.1 PRIMARY ENDPOINTS

- OS – defined as the time from the date of randomization to the date of death due to any cause in the ITT and PD-L1 positive analysis set

4.2 SECONDARY ENDPOINTS

- ORR – defined as the proportion of patients in the ITT and PD-L1 positive analysis set who had a complete response (CR) or partial response (PR) as assessed by the investigator per RECIST v1.1
- DoR – defined as the time from the first occurrence of a documented objective response to the time of relapse, as determined by the investigator per RECIST v1.1,

or death from any cause, whichever comes first in the ITT and PD-L1 positive analysis set

- PFS – defined as the time from the date of randomization to the date of the first objectively documented tumor progression as assessed by the investigator per RECIST v1.1 or death from any cause, whichever occurs first, in the ITT and PD-L1 positive analysis set

- HRQoL – measured using European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Lung Cancer (EORTC QLQ-LC13), Core 30 (EORTC QLQC30), and European Quality of Life 5-Dimensions, 5-level (EQ-5D-5L) scale

- Incidence and severity of treatment-emergent adverse events (TEAEs) graded according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE), v4.03

4.3 EXPLORATORY ENDPOINTS

- DCR – defined as the proportion of patients whose best overall response (BOR) is CR, PR or stable disease (SD) per RECIST v1.1

- CBR - defined as the proportion of patients who have CR, PR and SD that is ≥24 weeks in duration per RECIST v1.1

- PD-L1 expression, TMB, GEP, and tumor-infiltrating immune cells as predictive biomarkers for response

- Summary of serum concentrations of tislelizumab in the pharmacokinetic analysis set

- Assessments of immunogenicity of tislelizumab by determining the incidence of anti-drug antibodies (ADAs) in the ADA analysis set

5 SAMPLE SIZE CONSIDERATIONS

The original sample size calculation (ie, approximately 640 patients in China and Asia Pacific region (CAP) is based on the number of events required to demonstrate the OS superiority of Arm A to Arm B in ITT-CAP and ITT-CAP patients with PD-L1 positive tumors. At Protocol Amendment 1 of February 14, 2018, the sample size has been increased to include an additional 160 patients from rest of world (ROW), hence a total of approximately 800 patients will be recruited into the trial.

Six hundred and forty patients in ITT-CAP will be enrolled over a 16-month period at a constant enrollment rate and randomized in a 2:1 ratio to Arms A and B. The enrollment of 160 patients in ITT-ROW is expected to start approximately 8 months after that for the ITT-CAP and to last about 12 months. The median OS is assumed as 10 months in Arm B.

There is an approximately 87% power with planned 560 events to detect the difference between arms under the assumption of an OS HR (Arm A/Arm B) of 0.75 with a one-sided type I error of 0.02 in the ITT. An interim analysis is planned when approximately 426 deaths in the ITT analysis set have been observed, which represents 76% of the planned number of events in the ITT analysis set for the final analysis.

A Hwang-Shih-DeCani spending function with γ parameter of -2 based on the information fraction in the ITT analysis set is used in setting up the upper (efficacy) boundary. The stopping boundaries will be updated based on the actual death events observed in the ITT analysis set at the interim and final analyses.

The superiority test of OS in the PD-L1 positive analysis set will be performed only in the final analysis. Two hundred and seven deaths in the ITT patients with PD-L1 positive tumors are required to have approximately 86% power to detect an OS HR of 0.60 with a one-sided type I error of 0.007. Assuming the prevalence of PD-L1 positivity is 40% in the ITT analysis set, approximately 207 events in approximately 320 patients with PD-L1 positive tumors in the ITT analysis set are required.

The PD-L1 expression status will be closely monitored and enrollment of patients whose tumors are PD-L1 negative will be stopped as necessary through Interactive web response technology upon reaching ~60%, that is to ensure that the percentage of PD-L1 positive patients is no less than 40% of the ITT analysis set. The capping of PD-L1 negative patients to ~60% will be implemented in both ITT-CAP and ITT-ROW independently.

6 STATISTICAL METHODS

6.1 ANALYSIS SETS

The Intent-to-Treat (ITT) analysis set includes all randomized patients. Patients will be analyzed according to their randomized treatment arms. This will be the primary analysis set for efficacy analysis.

The PD-L1 positive analysis set (≥25% of TC with PD-L1 membrane staining) includes all randomized patients whose tumors were PD-L1 positive. Patients will be analyzed according to their randomized treatment arms. This will be the dual primary analysis population for efficacy analysis.

The Per-Protocol (PP) analysis set includes patients in the ITT analysis set who had no critical protocol deviations. Critical protocol deviations are a subset of major protocol deviations impacting primary efficacy analysis. Criteria for exclusion from the PP will be determined and documented before the database lock for the primary analysis.

The Safety (SAF) analysis set includes all randomized patients who received at least one dose of study drug. It will be the analysis set for the safety analyses. Patients who were randomized to tislelizumab arm but did not take any dose of tislelizumab will be included in the docetaxel arm

in the SAF. Patients who were randomized to docetaxel arm but took any dose of tislelizumab will be included in the tislelizumab arm in the SAF.

The HRQoL analysis set includes all randomized patients who received at least one dose of study drug and completed at least one HRQoL assessment. This will be the analysis set for HRQoL analysis.

The PD-L1 positive HRQoL analysis set includes all randomized patients whose tumors were PD-L1 positive and who received at least one dose of study drug and completed at least one HRQoL assessment. This will be the analysis set for PD-L1 positive HRQoL analysis.

The PK analysis set includes all patients who received at least 1 dose of tislelizumab per the protocol, for whom any post-baseline PK data are available.

The ADA analysis set includes all patients who received at least 1 dose of tislelizumab for whom both baseline ADA and at least 1 post-baseline ADA results are available.

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

Statistical programming and analyses will be performed using SAS® (SAS Institute, Inc., Cary, NC, USA), version 9.4 or higher, and/or other validated statistical software as required.

All descriptive statistics for continuous variables will be reported using mean, standard deviation (SD), median, 25 percentile (Q1), 75 percentile (Q3), minimum (Min), maximum (Max) and n. Categorical variables will be summarized as number (percentage) of patients.

The study Table Listing Graph shells will be provided in a separate document, which will show the content and format of all tables, listings, and graphs in detail.

6.2.1 Definitions and Computations

Baseline: Unless otherwise specified, a baseline value for ITT analysis set is defined as the last non-missing value collected before or at the time of randomization date, if not available, defined as last non-missing value collected before or at the time of first dose date. A baseline value for safety analysis set is defined as the last non-missing value collected before or at the time of first dose date. If the above-mentioned non-missing values were collected at both central labs and local labs, the non-missing values at central labs will be prioritized.

Unscheduled Visits: Unscheduled measurements will not be included in by-visit table summaries and graphs, but will contribute to best/worst case value where required (e.g. shift table). Listings will include scheduled and unscheduled data.

6.2.2 Conventions

Unless otherwise specified, the following conventions will be applied to all analyses:

- 1 year = 365.25 days. Number of years is calculated as (days/365.25) rounded up to 1

significant digit.

- 1 month = 30.4375 days. Number of months is calculated as (days/30.4375) rounded up to 1 significant digit.
- Age will be calculated as the integer part of (date of informed consent – date of birth + 1)/365.25
- P-values will be rounded to 4 decimal places. P-values that round to 0.0000 will be presented as '< 0.0001' and p-values that round to 1.000 will be presented as '> 0.9999'.

6.2.3 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in the SAP. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior or concomitant medications/procedures. Missing data for the HRQoL data will be handled according to each PRO instrument manual (Fayer & Machin, 2000 in The EORTC QLQ-C30 (Third Edition), 2001; https://euroqol.org/publications/user-guides/). Every effort should be made to ensure complete death dates. In the rare case, if day of death date is missing, death date will be imputed as the max of last available date showing patients alive + 1 and first day of year/month of death date. However, death date will not be imputed if year or month is missing.

By-visit endpoints will be analyzed using observed data, unless otherwise specified. For observed data analyses, missing data will not be imputed and only the observed records will be included.

6.2.4 Adjustment for Covariates

If any baseline characteristics are unbalanced, certain analysis adjusted for covariates may be considered.

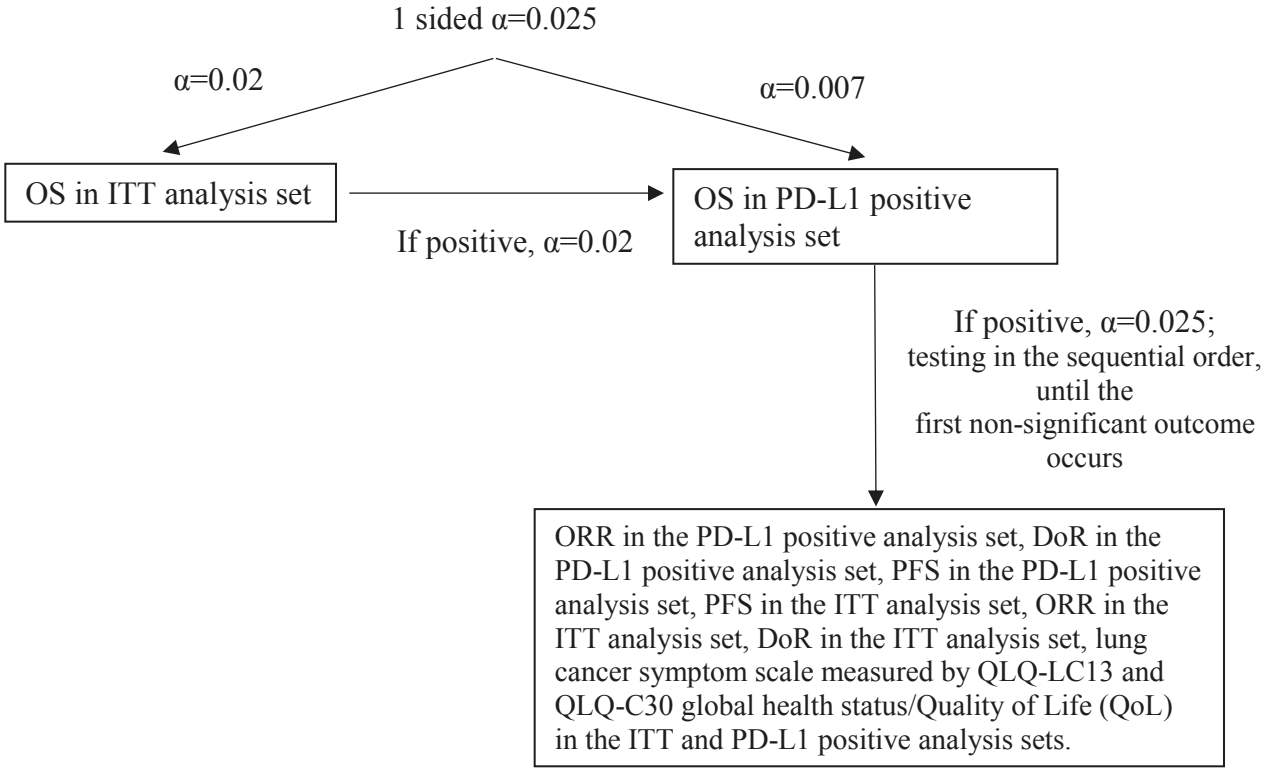
6.2.5 Multiplicity Adjustment

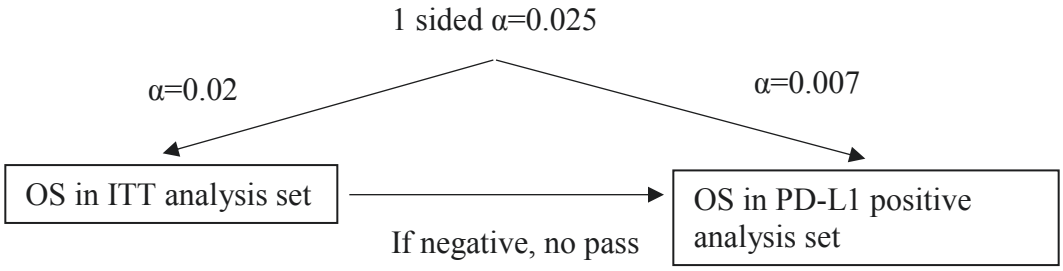
A sequential testing with alpha splitting approach will be implemented in the following orders to control overall type I error of 0.025 for the primary endpoints (OS in ITT and PD-L1 positive subset). The α allocation accounts for the positive correlation between the test statistics in the 2 Analysis Sets (ie, PD-L1 positive is a subset of the ITT Analysis Set). The overall type I error is controlled at 0.025 when at least 30% of the deaths in the ITT Analysis Set are from the PD-L1 positive subset. The α of 0.007 in the PD-L1 testing will be adjusted downwards if the final observed percentage is lower.

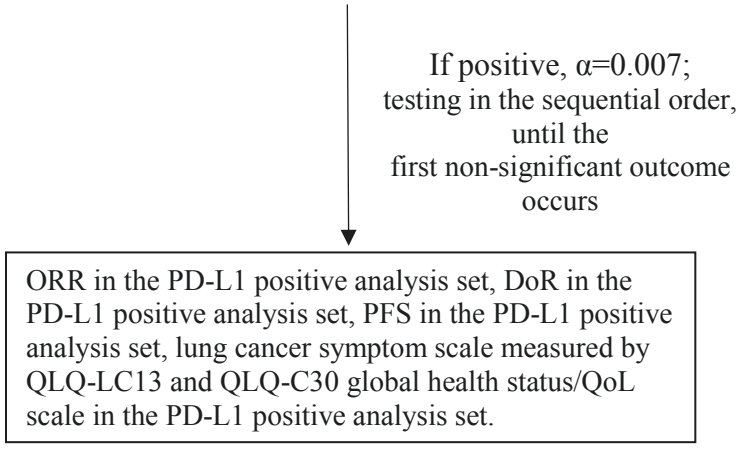
- OS in ITT with α of 0.02 (0.0112 for interim analysis, 0.0153 for final analysis)
- If the OS hypothesis in ITT is rejected,
    - OS in PD-L1 positive analysis set

- testing in the sequential order of ORR in the PD-L1 positive analysis set, DoR in the PD-L1 positive analysis set, PFS in the PD-L1 positive analysis set, PFS in the ITT analysis set, ORR in the ITT analysis set, DoR in the ITT analysis set, lung cancer symptom scale measured by QLQ-LC13 and QLQ-C30 global health status/Quality of Life (QoL) in the ITT and PD-L1 positive analysis sets.

- If the OS hypothesis in the ITT analysis set cannot be rejected, hypothesis testing will be carried out sequentially only in the PD-L1 positive Analysis Set for OS, ORR, DoR, PFS, lung cancer symptom scale measured by QLQ-LC13 and QLQ-C30 global health status/QoL scale at α of 0.007.







6.3 PATIENT CHARACTERISTICS

6.3.1 Patient Disposition

The number and percentage of patients randomized, treated, permanently discontinued from study treatment, remained on treatment, discontinued from study, and remained in study will be summarized in the ITT analysis set. The primary reasons for study treatment discontinuation and study discontinuation will be summarized according to the categories in the CRF. Study follow-up time and primary reason for screen failure will be summarized.

6.3.2 Protocol Deviations

Critical protocol deviations are a subset of major protocol deviations impacting primary efficacy analysis. Criteria for major protocol deviations and critical protocol deviations impacting primary efficacy will be established and patients with major and critical protocol deviations will be identified and documented before the database lock.

Both major protocol deviations and critical protocol deviations, including relatedness to COVID-19, will be summarized and listed by category in the ITT analysis set.

6.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized in the ITT analysis set using descriptive statistics.

Continuous demographic and baseline variables include age, BMI (in kg/m^2), body weight (in kg), baseline target lesions sum of diameters by investigator (in mm), time from initial diagnosis to study entry; categorical variables include sex, age group (<65 years, ≥65 - <75 years, ≥75 - <85 years, ≥85 years), race, ethnicity, country, region (China, ROW), ECOG performance status at baseline, smoking status, EGFR mutation, ALK rearrangement, disease stage, brain metastasis, liver metastasis and location of distant metastases.

Stratification factors including histology, line of therapy, and PD-L1 expression will be summarized in the ITT analysis set.

Cancer associated symptoms at baseline will also be summarized by SOC, preferred term and CTCAE grade.

6.3.4 Prior Anti-Cancer Systemic Therapies, Radiotherapy and Surgeries

The number and percentage of patients with any prior anti-cancer systemic therapies, time from end of last therapy to study entry, type of prior anti-cancer systemic therapy and setting of prior anti-cancer systemic therapy will be summarized in the ITT analysis set.

The number and percentage of patients with any prior anti-cancer radiotherapy, and time from end of last radiotherapy to study entry will be summarized in the ITT analysis set.

The number and percentage of patients with any prior anti-cancer surgeries, intention of surgery and time from last surgery to study entry will be summarized in the ITT analysis set.

6.3.5 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the version of World Health Organization Drug Dictionary (WHO DD) drug codes currently in effect at BeiGene at the time of database lock, and will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

The number and percentage of patients reporting prior and concomitant medications will be summarized respectively by ATC medication class and WHO DD preferred term (PT) in the safety analysis set. Prior medications will be defined as medications that received within 30 days before randomization and stopped before the first dose of study drug. Concomitant medications will be defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 30 days after the patient's last dose (as of Safety Follow-up Visit). In addition, telephone contacts with patients should be conducted to assess irAEs and associated concomitant medications at 60 and 90 days (± 14 days) after the last dose of tislelizumab regardless of whether or not the patient starts a new anticancer therapy.

6.3.6 Medical History

Medical history will be coded using MedDRA version currently in effect at BeiGene at the time of database lock. The number and percentage of patients reporting a history of any medical condition, as recorded on the eCRF, will be summarized by System Organ Class (SOC) and Preferred Term (PT) in the ITT analysis set.

6.4 EFFICACY ANALYSIS

If not specified otherwise, efficacy analyses described in this section will be based on the ITT analysis set and the PD-L1 positive analysis set, respectively.

6.4.1 Primary Efficacy Endpoints

Overall Survival

Overall Survival in the ITT Analysis Set:

OS is defined as the time from randomization to death from any cause. Data for patients who are not reported as having died at the time of analysis will be censored at the date last known to be alive. Data for patients who do not have post-baseline information will be censored at the date of randomization.

OS will be compared between tislelizumab (Arm A) and docetaxel (Arm B) in a stratified log-rank test using a significance level of 0.02 (one-sided).

The null hypothesis to be tested is:

H_0: OS in Arm A = OS in Arm B

against the alternative hypothesis:

H_a: OS in Arm A ≠ OS in Arm B

This will be the primary analysis once the targeted numbers of deaths are reached in the ITT analysis set.

The p-value will be calculated from a stratified log-rank test at one-sided significance level α=0.02 based on the stratification factors defined in Section 2 using the actual values from EDC (histology and line of therapy) and from central lab (PD-L1 expression).

Kaplan-Meier method will be used to estimate median and other quartiles of OS along with its 95% confidence interval (constructed using Brookmeyer and Crowley method). Kaplan-Meier survival probabilities for each arm will be plotted over time. Event free rate at selected timepoints will be estimated with 95% confidence interval using Greenwood formula. Follow-up time will be estimated by the reverse Kaplan-Meier method.

The hazard ratio for OS between tislelizumab and docetaxel ($HR_{A/B}$) will be estimated using a stratified Cox proportional hazard model with Efron's method of tie handling, with treatment arm as a factor and stratified by the same stratification variables used for the stratified log-rank test. The 95% CI for the HR will be provided. Unstratified log-rank test and Cox regression model will also be performed to provide p-value, HR and corresponding 95% CI.

In order to evaluate the robustness of the OS, several sensitivity analyses are planned.

The sensitivity analysis 1 is the same as the primary analysis except that it will be based on the stratification factors using the values from Interactive Response Technology, by which patients were randomized.

The sensitivity analysis 2 is the same as the primary analysis except that it will use Rank Preserving Structural Failure Time Model (RPSFTM) to adjust survival estimates in the presence of arm B patients receiving any subsequent immunotherapy after discontinuation of docetaxel.

The sensitivity analysis 3 is the same as the primary analysis except that a patient will be censored at the date last known to be alive before his/her COVID-19 related drug administration protocol deviation.

When there are over 10% ITT patients who had critical protocol deviations, the sensitivity analysis 4 in the PP analysis set will be implemented in the same way as the primary analysis.

Overall Survival in the PD-L1 Positive Analysis Set:

The hypothesis testing of OS in the PD-L1 positive analysis set will be carried out at a significance level of 0.007. If the OS hypothesis in the ITT analysis set is rejected, its corresponding α will be shifted to the testing in the PD-L1 positive analysis set (ie, a total α of 0.025). Similar statistical methods as described above will be applied with histology and line of therapy as strata in the stratified analyses.

Subgroup Analyses

To determine if the treatment effect is consistent across various subgroups, the HR estimates of OS from an unstratified Cox model and its 95% CI will be estimated and plotted within each category of the following variables (a subgroup may not be analyzed if it includes <10% of the ITT analysis set):

- PD-L1 expression in TC (≥25% TC versus <25% TC)
- histology (squamous versus non-squamous)
- line of therapy (2 versus 3)
- age (<65 versus ≥65 years)
- sex (female versus male)
- ECOG PS (0 versus 1)
- smoking status (never versus former/current)
- race (Asian, White, Other)
- region (China, Europe, Other)
- brain metastasis at baseline (yes versus no)

- liver metastasis at baseline (yes versus no)
- disease stage (locally advanced versus metastatic)
- EGFR mutation at baseline (wild type versus unknown)
- ALK rearrangement at baseline (wild type versus unknown)

Forest plot of subgroup analysis in OS will be provided. Additional subgroup analysis may also be conducted per additional prognosis factors as suggested. Subgroup analysis of secondary endpoints may also be conducted.

6.4.2 Secondary Efficacy Endpoints

Objective Response Rate

ORR is the proportion of patients who had a CR or PR as assessed by the investigator per RECIST v1.1 in the ITT and PD-L1 positive analysis set. Patients without any postbaseline assessment will be considered non-responders. Patients without measurable disease at baseline will also be considered as non-responders. The difference in ORR between arms will be evaluated using the Cochran-Mantel-Haenszel (CMH) chi-square test with the actual stratification factors as strata. The two-sided 95% CIs for the odds ratio and the difference in ORR will be calculated, as well as Clopper-Pearson 95% CIs for the ORR within each arm.

In addition, the number and percentage of patients for each of the BOR categories will be presented.

A waterfall plot of best percent change in sum of target lesion diameters from baseline will be provided by treatment arm. The patients in each arm will be ordered by the percentage, patients with the largest percentage will be presented on the right.

Progression-Free Survival

PFS is defined as the time from randomization to the first objectively documented disease progression as assessed by the investigator per RECIST v1.1 or death from any cause, whichever occurs first, in the ITT and PD-L1 positive analysis sets. The actual tumor assessment visit date will be used to calculate PFS. The PFS censoring rules are specified in Appendix 10.1. Similar methodology except for sensitivity analyses used to evaluate OS will be applied to analysis of PFS.

Duration of Response

DoR is defined for patients with an objective response as the time from the first documented objective response to documented disease progression as assessed by the investigator using RECIST v1.1, or death from any cause, whichever occurs first, in the ITT and PD-L1 positive analysis sets. Only the subset of patients who show a complete response or partial response will

be included in the DoR analysis. Data for patients who are alive and who have not experienced disease progression at the time of analysis will be censored at the date of the last tumor assessment. If no tumor assessments were performed after the date of the first occurrence of the objective response (CR or PR), DoR will be censored at the date of the first occurrence of the objective response. Median DoR and corresponding 95% CIs will be estimated using Kaplan-Meier methodology for each treatment arm. Comparisons of DoR between treatment arms will be made using the log-rank test.

Health-Related Quality of Life

Analysis Method

Descriptive statistics will be used for all HRQoL analyses. All HRQoL analyses will be in the HRQoL analysis set and/or the PD-L1 positive HRQoL analysis set, unless otherwise specified. Both EORTC QLQ-C30 and EORTC QLQ-LC13 instruments have been extensively tested for reliability and validity (Bergman et al, 1994; Osoba et al, 1994; Groenvold et al, 1997).

A linear mixed-effect model for repeated measures (MMRM) will be used to compare between the 2 treatment arms. The model includes the repeated measures (including baseline) of the index scores of QLQ-C30 and LC13 and dyspnoea, coughing and pain in chest of LC13 as the dependent variables, and treatment arm, randomization stratification factors, intercept and slope of time as fixed effects. The random subject effects include subject random intercept and subject random slope of time. Interaction between treatment and time since randomization or quadratic term of time since randomization may be considered to be included in the final model if it is significant when it needs to be formally tested at the final analysis. They will be performed in the ITT analysis set and the PD-L1 positive analysis set, respectively. Between-group comparisons will be reported as differences with the 95% CI and p-value.

In addition, changes from baseline in global health status/QoL of QLQ-C30 and the functional and symptom scales of QLQ-C30, QLQ-LC13 scales and EQ-5D-5L (descriptive scores and visual analog scales) will be summarized descriptively.

Compliance

Compliance for the EORTC QLQ-C30, EORTC QLQ-LC13 and EQ-5D-5L modules, defined as the proportion of questionnaires actual received out of the expected number (i.e, number of patients on treatment), in the HRQoL analysis set will be summarized for each assessment time point and treatment arm.

Change from Baseline by Visit

For each scale or item of EORTC QLQ-C30, EORTC QLQ-LC13 and EQ-5D-5L, summary statistics at each assessment time point and change from baseline will be presented by treatment

arm in tables. Boxplot depicting the mean scores over time of global health status/quality of life will be provided for each treatment arm.

Details of HRQoL scoring are specified in Appendix 10.2 according to the algorithm described in the EORTC QLQ-C30 and EORTC QLQ-LC13 scoring manual (Fayers 2001) and EQ5D-5L (https://euroqol.org/publications/user-guides/).

Time to Deterioration (TTD) of HRQoL

Time to deterioration (TTD) is defined as the time from randomization to first occurrence of a worsening score confirmed at the subsequent visit or death from any cause. The minimum important clinically meaningful difference change (e.g. worsening) in symptoms of QLQ-C30 and QLQ-LC13 is defined as ≥10 points increase from baseline (Osoba et al 1998; King, 1996; Maringwa et al 2011). The clinically meaningful deterioration in function and global health status/quality of life is defined as ≥10 points decrease from baseline. The median TTD of QLQ-C30 global health status/quality of life and index score of the QLQ-LC13 will be calculated using Kaplan-Meier estimates, and presented with 2-sided 95% CIs.

EQ-5D-5L

Five level response to EQ-5D-5L will be summarized as a categorical variable by tabulating frequency of each response category by visit. The self-rated health state scale at each visit and its change from baseline will be summarized in descriptive statistics (n, mean, standard deviation, median, minimum, maximum).

6.4.3 Exploratory Efficacy Endpoints

Disease Control Rate per the Investigator

DCR is defined as the proportion of patients with objective response (CR or PR), Non-CR/Non-PD, or stable disease maintained for ≥ 9 weeks (with allowable visit window) using RECIST v1.1. DCR per the investigator will be analyzed. Similar methodologies for analysis of ORR will be applied.

Clinical Benefit Rate per the Investigator

CBR is defined as the proportion of patients who have CR, PR, Non-CR/Non-PD and SD that is ≥24 weeks in duration per RECIST v1.1. CBR per the investigator will be analyzed. Similar methodologies for analysis of ORR will be applied.

Time to Response per the Investigator

TTR per the investigator is defined for patients with an objective response as determined by the investigator as the time from randomization to the first occurrence of a CR or PR as assessed by the investigator using RECIST v1.1. Only the subset of patients who show a CR or PR will be

included in the TTR analysis. TTR will be summarized for descriptive purposes. The mean, SD, median, and range of TTR will be provided.

Time to First Subsequent Anti-cancer Systemic Therapy (TFST)

TFST is defined for patients with the use of subsequent anti-cancer systemic therapy as the time from end of study treatment to first dose of subsequent anti-cancer systemic therapy. The mean, SD, median, and range of TFST will be provided.

Subsequent Anti-cancer Therapy

Subsequent anti-cancer therapy will be summarized by percentage, category and preferred term (PT) in the ITT and PD-L1 positive analysis sets for each treatment arm.

PD-L1 Expression as a Predictive Biomarker for Response

Distribution of PD-L1 expression will be examined in the ITT analysis set. Association between PD-L1 expression (not restricted to the pre-specified cut off level of 25%) and tislelizumab treatment effect over docetaxel (OS, ORR, PFS, DoR, DCR, CBR) will be explored. Exploratory analyses may not be included in the CSR for this study.

6.4.4 Efficacy Endpoints in China patients

According to the feedback from Center for Drug Evaluation (CDE) of National Medical Products Administration (NMPA) to Protocol Amendment 3 on July 28, 2020, it was required that China patients to be used as primary efficacy analysis population for China supplementary Biologics License Application (sBLA) using the O'Brien-Fleming boundary defined in the original protocol, and ITT population as supporting analysis results. Therefore, analyses of efficacy endpoints such as OS, PFS, ORR and DoR in China patients will be performed as well.

6.5 SAFETY ANALYSES

All safety analyses will be performed by treatment arm based on the safety analysis set. The incidence of treatment-emergent adverse events (TEAEs) and SAEs will be summarized. Laboratory test results, vital signs, ECG, ECOG and their changes from baseline will be summarized using descriptive statistics (e.g., n, mean, standard deviation, median, Q1, Q3, minimum, maximum for continuous variables; n [%] for categorical variables). Abnormal values will be flagged.

6.5.1 Extent of Exposure

Extent of exposure to each study drug will be summarized descriptively by the following variables:

- Number of treatment cycles
- Duration of exposure (weeks) is defined as:

(date of last dose of study drug + 21 days – date of first dose of study drug)/7, with censored by death date and cutoff date, without censoring when calculating actual dose intensity.

- Cumulative dose (mg for tislelizumab and mg/m^2 for docetaxel): the sum of all actual dose of study drug, given from first to last administration

- Actual dose intensity (ADI) (mg/cycle for tislelizumab and mg/m^2/cycle for docetaxel) is defined as

  Cumulative dose (mg for tislelizumab and mg/m^2 for docetaxel) received by a patient / Duration of exposure (cycles)

- Relative dose intensity (RDI) in % is defined as:

$$100 \times \frac{\text{ADI (mg/cycle for tislelizumab and mg/m}^2\text{/cycle for docetaxel)}}{\text{Planned Dose Intensity (mg/cycle for tislelizumab and mg/m}^2\text{/cycle for docetaxel)}}$$

  where planned dose intensity of tislelizumab equals to 200 mg/cycle and planned dose intensity of docetaxel equals to 75 mg/m^2/cycle.

6.5.2 Adverse Events

AEs will be graded by the investigators using CTCAE v4.03. The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to the MedDRA (Version 23.0 or higher) lower level term closest to the verbatim term. The linked MedDRA preferred term (PT) and primary system organ class (SOC) are also captured in the database.

A treatment-emergent AE (TEAE) is defined as an AE that had an onset date or a worsening in severity from baseline (pretreatment) on or after the first dose of either study drug up to 30 days following study drug discontinuation or initiation of new anticancer therapy, whichever occurs first. TEAE classification also applies to irAEs and drug-related SAEs recorded up to 90 days after discontinuation from tislelizumab, regardless of whether or not the patient starts a new anticancer therapy. Only those AEs that were treatment-emergent will be included in summary tables. All AEs, treatment emergent or otherwise, will be presented in patient data listings.

An overview table of patients with at least one TEAE will be presented with the incidence of:

- patients with any TEAEs

- patients with any TEAEs with grade ≥ 3

- patients with any serious TEAEs

- patients with any TEAEs leading to death
- patients with any TEAEs leading to permanent treatment discontinuation
- patients with any TEAEs leading to treatment modification
- patients with any treatment-related TEAEs
- patients with any treatment-related TEAEs and grade ≥ 3
- patients with any treatment-related serious TEAEs
- patients with any treatment-related TEAEs leading to death
- patients with any treatment-related TEAEs leading to permanent treatment discontinuation
- patients with any irAEs

Treatment-related TEAEs include those events considered by the investigator to be related or with missing assessment of the causal relationship. For patients with multiple occurrences of the same event will be counted only once in the summary tables if otherwise specified, and the maximum grade per CTCAE v4.03 will be used.

If the grade is missing for one of the treatment-emergent occurrences of an adverse event, the maximal severity on the remaining occurrences with the same preferred term of the same patient will be used. If the patient has no other TEAEs with the same preferred term, then impute as the maximal severity on all TEAEs with the same preferred term; if the severity is missing for all the occurrences, do not impute, a "missing" category will be added in the summary table.

The incidence of following TEAEs will be reported by SOC and PT, sorted by decreasing frequency of the SOC and PT:

- TEAEs by maximum grade
- TEAEs with grade ≥ 3
- Serious TEAEs
- TEAEs leading to death
- TEAEs leading to permanent treatment discontinuation
- Treatment-related TEAEs

- Treatment-related TEAEs and grade ≥ 3
- Treatment-related serious TEAEs
- Treatment-related TEAEs leading to death

All deaths and causes of death will be summarized by treatment arm, including those occurred during the study treatment period and those reported during the survival follow-up period after treatment discontinuation.

Additionally, if exposure to treatment is very different between the two arms, exposure-adjusted event rate (EAER) per 100 person-months may be reported by treatment arm. The EAER per 100 person-months is defined as 100 times the number of events divided by the total exposure time (in months) among patients included in the analysis. Patients with multiple occurrences of the event will be counted multiple times in the numerator. The exposure time is the treatment duration for all patients with/without the event. The total exposure time in months is calculated by dividing the sum of exposure time in days over all the patients included in the analysis by 30.4375. The EAER per 100 person-months is interpreted as the expected number of events per 100 person-months of exposure to the investigational products.

TEAEs will also be summarized and presented for certain subgroups.

6.5.3 Laboratory Values

Descriptive summary statistics (n, mean, standard deviation, median, minimum, maximum for continuous variables; n [%] for categorical variables) for selected laboratory parameters described in Table 1 and their changes from baseline will be summarized by visit.

Selected laboratory parameters that are graded in NCI-CTCAE v4.03 will be summarized by shift from baseline CTCAE grades to maximum post-baseline grades. In the summary of laboratory parameters by CTCAE grade, parameters with CTCAE grading in both high and low directions will be summarized separately.

The patients who met Hy's Law by lab value will be summarized.

Patient data listings of selected laboratory parameters will be provided.

Table 1. Clinical Laboratory Tests

Serum Chemistry	Hematology	Thyroid Function
ALP Alanine aminotransferase (ALT) Aspartate aminotransferase (AST) Total bilirubin Creatinine	Hemoglobin White blood cell (WBC) count Neutrophil (Absolute) Lymphocytes (Absolute) Platelet count	Free Triiodothyronine (FT3) Free Thyroxine (FT4) Thyroid Stimulating Hormone (TSH)

Potassium Sodium Calcium (corrected) Creatine kinase (CK) Creatine kinase-cardiac muscle isoenzyme (CK-MB) Lactate dehydrogenase Glucose		

6.5.4 Vital Signs

Descriptive statistics for vital sign parameters (i.e., systolic and diastolic blood pressure, pulse rate, body temperature, and weight) and changes from baseline will be summarized by visit.

6.5.5 Ophthalmologic Examination

Ophthalmologic examination results (optical coherence test and equivalent exams normal/abnormal) will be summarized.

6.5.6 Electrocardiograms (ECG)

12-Lead ECG will be performed during baseline and multiple time post-baseline points (refer the time points to the protocol study assessments and procedures schedule). Overall evaluation and clinical significance will be summarized.

6.5.7 ECOG

A shift table from baseline to worst post-baseline in ECOG performance score will be summarized.

6.6 PHARMACOKINETIC ANALYSES

PK samples will be collected only in patients randomized to receive tislelizumab. Tislelizumab post-dose and Ctrough (pre-dose) will be tabulated and summarized for each cycle at which these concentrations are collected. Descriptive statistics will include means, medians, ranges, standard deviations, coefficient of variation (CV%), geometric mean and geometric CV%, as appropriate.

Additional PK analyses, including population PK analyses and exposure-response (efficacy, safety endpoints) analyses may be conducted as appropriate and the results from these analyses will be reported separately from the CSR.

6.7 IMMUNOGENIC ANALYSIS

Human anti-drug antibodies (ADA) to tislelizumab will be assessed during the study as defined in the protocol.

ADA attributes:

- Treatment boosted ADA is defined as ADA positive at baseline that was boosted to a 4-fold or higher level following drug administration.

- Treatment-induced ADA is defined as ADA negative at baseline and ADA positive post-baseline.

- Persistent ADA response is defined as Treatment-induced ADA detected at 2 or more time points during treatment or follow-up, where the first and last ADA positive samples are separated by 16 weeks or longer; or detected in the last time point.

- Transient ADA response is defined as treatment-induced response that is not considered persistent.

- Neutralizing ADA is defined as ADA that inhibits or reduces the pharmacological activity.

ADA response endpoints:

- ADA incidence is defined as sum of treatment-emergent ADA, which include both treatment-induced and treatment-boosted ADA-positive patients, as a proportion of the ADA evaluable population.

- ADA prevalence is defined as proportion of all patients that are ADA positive, including pre-existing ADA, at any time point.

The immunogenicity results will be summarized using descriptive statistics by the number and percentage of patients who develop detectable ADA. The incidence of positive ADA and neutralizing ADA will be reported for evaluable patients. The effect of immunogenicity on PK, efficacy, and safety may be evaluated if data allow.

6.8 OTHER EXPLORATORY ANALYSES

Other potential predictive markers including but not limited to GEP, TMB, and tumor-infiltrating immune cells will be assessed and presented in a separate report.

6.9 ANALYSES TO EVALUATE COVID-19 IMPACT

Summary of COVID-19 impact will be presented as tables or listings for patient disposition, protocol deviations, sensitivity analysis of OS, tumor assessment, AEs and laboratory parameters.

7 INTERIM ANALYSIS

There will be one interim analysis of OS performed in the ITT analysis set. The interim analysis will be performed when approximately 426 deaths (76% of the target number of 560 deaths) among the 2 treatment arms are observed in the ITT analysis set. It is estimated that it will take

approximately 23.1 months to observe 426 events. The final analysis of OS will take place after 560 deaths are observed in the ITT analysis set and 207 deaths are observed in its subgroup of patients with PD-L1 positive tumors. Thus, the predefined number of deaths in the ITT analysis set will trigger the interim and final analyses. The information fraction used in α spending function will be based on the observed number of deaths in the ITT analysis set at the corresponding time points. A Hwang-Shih-DeCani spending function with γ parameter of -2 will be used in setting up the upper (efficacy) boundary. Stopping boundaries (p-value and Z score) of superiority test for OS at the interim and final analyses in the ITT analysis set, as well as OS at the final analysis in the PD-L1 positive analysis set are shown in Table 2. The boundaries for hypothesis testing in OS will be updated according to the actual numbers of death events in the interim and final analyses, using the pre-specified α spending function.

Table 2. Stopping Boundaries (p-value and Z score) and Approximate HR Threshold of Interim and Final Analyses of OS

	Time (months)	# Deaths	p-value (Z score) for Efficacy	Approximate HR Threshold for Efficacy
Interim analysis in ITT	23.1	426	<0.0112 (>2.28)	<0.791
Final analysis in ITT	31.0	560	<0.0153 (>2.16)	<0.824
Final analysis in PD-L1 positive	31.0	207	<0.007 (>2.46)	<0.696

Abbreviations: HR = hazard ratio; ITT = intent-to-treat (analysis set); PD-L1 – programmed cell death protein ligand 1

Additionally, as required by CDE, at the time of interim analysis in the ITT, the analysis of OS among China patients will be performed as well. The upper (efficacy) boundary is based on O'Brien-Fleming boundary approximated by Hwang-Shih-DeCani spending function with γ parameter set at -4. Stopping boundaries (p-value and Z score) of superiority test for OS at the interim and final analyses are shown in the table below as reference to facilitate the review of CDE. Meanwhile, the boundaries presented in the table below will be updated from the actual death events observed at the time of the interim analysis and final planned number of deaths in China patients, using the spending function.

	Time (months)	# Deaths	p-value (Z score) for Efficacy	Approximate HR Threshold for Efficacy
Interim analysis in China patients	23.1	372	<0.0088 (>2.37)	<0.770
Final analysis in China patients	31.0	465	<0.0176 (>2.11)	<0.813
Final analysis in PD-L1 positive China patients	31.0	172	<0.007 (>2.46)	<0.672

Abbreviations: HR = hazard ratio; PD-L1 – programmed cell death protein ligand 1

8 CHANGES IN THE PLANNED ANALYSIS

Not applicable.

9 REFERENCES

Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M (1994) The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. Eur J Cancer 30A: 635–642.

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

Clopper, C. and Pearson, ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika 1934; 26: 404-413.

Collett D. Modelling survival data in medical research. New York: Chapman & Hall; 1994.

Common Terminology Criteria for Adverse Events (CTCAE). Version 4.0. United States Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, USA, May 28, 2009.

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

Fayers PM, Aaronson NK, Bjordal K,Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. Brussels: EORTC, EORTC QLQ-C30 Scoring Manual (3rd edition). 2001. ISBN: 2-9300 64-22-6.

Food and Drug Administration Center for Drug Evaluation Research CDER and Center for Biologics Evaluation and Research. FDA Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, 2007.

Greenwood M. “The natural duration of cancer”. Reports on Public Health and Medical Patients (London: Her Majesty’s Stationery Office). 1926; 33:1-26.

Groenvold M, Klee MC, Sprangers MA, Aaronson NK (1997) Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 50: 441–450.

Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Washington, DC: United States Food and Drug Administration; 2007.

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

King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res 1996; 5: 555–567.

Maringwa J, Quinten C, King M, et al. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. Annals of Oncology 2011; 22: 2107–2112.

Osoba D, Rodrigues G, Myles J, Zee B, Pater J (1998) Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 16: 139–144.

Osoba D, Zee B, Pater J, Warr D, Kaizer L, Latreille J (1994) Psychometric properties and responsiveness of the EORTC quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. Qual Life Res 3: 353–364.

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun Stat-Theor M 1991;20(8):2609-31.

10 APPENDIX

10.1 PFS CENSORING RULES

Definition of Progression Date: Progression date is assigned to the first time when tumor progression was documented.

The PFS derivation rules in this SAP follow the Table C1 and C2 described in Appendix C of Food and Drug Administration (FDA) "Guidance for Industry Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics (2015)", which includes documented progression only.

Censoring rules for primary analysis of PFS are summarized in Table 3.

Table 3. Censoring Rules for Primary Analysis of PFS Per RECIST version 1.1

No.	Situation	Primary Analysis
1	Incomplete or no baseline tumor assessments	Censored at randomization date
2	No postbaseline tumor assessment and no death	Censored at randomization date
3	No postbaseline tumor assessment and death	Died at date of death
4	Progression documented between scheduled visits	Progressed at date of documented progression
5	No progression	Censored at date of last adequate tumor assessment with no documented progression
6	New anticancer treatment started	Censored at date of last adequate tumor assessment before date of new anticancer treatment
7	Death between adequate assessment visits	Died at date of death
8	Death or progression after ≥2 missed visit	Censored at date of last adequate tumor assessment prior to the ≥2 missed tumor assessments

10.2 HEALTH RELATED QUALITY OF LIFE

EORTC QLQ-C30 measures HRQoL of general cancer and includes two items that measure Global health status and quality of life (Global health status/QoL). The instrument is also consisted of two Functional Scales and Symptom Scales/ Items. Functional scales are Physical (2 items), Role (2 items), Emotional (4 items), Cognitive (2 items) and Social (2 items) functioning. Symptom scales are Fatigue (3 items), Nausea (2 items), Pain (2 items) symptoms. The single-items measuring symptoms include Dyspnoea, Insomnia, Appetite loss, Constipation, Diarrhoea and Financial Difficulties.

QLQ-LC13, which is the lung cancer module of the QLQ-C30, measure symptoms specific to lung cancer and its treatment. It is comprises of 13 items including Dyspnoea scale (2 items) and single items measuring Coughing, Haemoptysis, Sore month, Dysphagia, Perpherial neuropathy, Alopecia and Pain in chest, Pain arm/shoulder and Pain other parts.

EQ5D-5L measures general HRQoL consists for 5 scales (descriptive dimension) and a Visual Analogue Scale (VAS). The descriptive Dimensions scale includes Mobility, Self-Care, Usual Activities, Pain/Discomfort and Anxiety/Depression.

QLQ-C30 and QLQ-LC13 scale scores as well as EQ-5D-5L will be calculated as described below.

Scoring Process

QLQ-C30 and QLQ-LC13: The principle for scoring applies to all scales/scores: Raw scores are calculated as the average of the items that contribute to the scale.

A linear transformation to standardize the raw scores is utilized, so that the scores are ranged from 0 to 100. Increases in scores for functional domains (e.g., physical, role, social, emotional, etc.) are improvements while increases in scores for symptoms (e.g., fatigue, vomiting and nausea, diarrhea, pain, etc.) are deteriorations.

Missing Items
If at least half of the items for a scale are answered, then all the completed items are used to calculate the score. Otherwise, the scale score is set to missing.

In practical terms, if items I_1, I_2, ... I_n are included in a scale, the procedure is as follows:

Raw Score
For all scores, the raw score (RS), is the mean of the component items
RS = $(I_1+I_2+\ldots+I_n)$/n

Derived Scale
The derived scales are obtained from the raw scores as defined in the EORTC manual. The derived scales have a more intuitive interpretation: larger function scale or global health status /
QoL are improvements while larger symptom scales (e.g., pain, nausea, etc.) are deteriorations.

The derivation formulas are as follows.

Linear transformation

Apply the linear transformation to 0-100 to obtain the score S,

Functional scales: $S = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$

Symptom scales / items: $S = \{(RS-1)/range\} \times 100$

Global health status / QoL: $S = \{(RS-1)/range\} \times 100$

The Index scores

To calculate QlQ-C30 index-score, individual functioning scale are subtracted by 100 to convert them into having the same meaning as symptom/problem scales. These 6 subtracted scores are subsequently summed with the 9 symptom/problem scales, and then divided by 15 (the total number of QLQ-C30 scales). A higher C30 index-score reflects a worse overall HRQOL. This is the mathematical formula:

QlQ- C30 index score = ∑[(100 - Physical functioning score), (100 - Role functioning score), (100 - Emotional functioning score), (100 - Cognitive functioning score), (100 - Social functioning score) , (100 - global QOL score), Fatigue score, Nausea/vomiting score , Pain score, Dyspnoea score, Insomnia score, Appetite loss score, Constipation score, Diarrhea score, Financial Difficulty score] ÷ 15

LC13 index-score is defined as the sum of all 10 QLQ-LC13 symptom/problem scales divided by 10 (the total number of QLQ-LC13 scales). A higher LC13 index-score reflects a worse overall HRQOL. This is the mathematical formula:

LC13 index score = Σ (scores of Dyspnoea, Coughing, Haemoptysis, Sore mouth, Dysphagia, Peripheral neuropathy, Alopecia, Pain in chest, Pain in arm or shoulder, Pain in other parts) ÷ 10

Table 4 Scoring of QLQ-C30

	Scale	Number of items	Item range	Item Numbers
Global health status/ QoL Global health status/QOL	QL2	2	6	29,30
Functional Scales				
Physical functioning	PF2	5	3	1, 2, 3, 4, 5

Role functioning	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21, 22, 23, 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
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

Table 5 Scoring of QLQ-LC13

	Scale	Number of items	Item range	Item Numbers
Symptom scales/items				
Dyspnoea	LCDY	3	3	3,4,5
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

EQ5d-5L: for the 5 level Dementions, scores on a 5-point likert scale of 1 to 5, with level 1 indicating no problem and level 5 indicating extreme problems.

VAS includes a scale of 0 to 100 with higher scores indicating better health status.

Table 6 Scoring of Eq5D-5L

MOBILITY	
I have no problems in walking about	☐
I have slight problems in walking about	☐
I have moderate problems in walking about	☐
I have severe problems in walking about	☐
I am unable to walk about	☐
SELF-CARE	
I have no problems washing or dressing myself	☐
I have slight problems washing or dressing myself	☐
I have moderate problems washing or dressing myself	☐
I have severe problems washing or dressing myself	☐
I am unable to wash or dress myself	☐
USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities)	
I have no problems doing my usual activities	☐
I have slight problems doing my usual activities	☐
I have moderate problems doing my usual activities	☐
I have severe problems doing my usual activities	☐
I am unable to do my usual activities	☐
PAIN / DISCOMFORT	
I have no pain or discomfort	☐
I have slight pain or discomfort	☐
I have moderate pain or discomfort	☐
I have severe pain or discomfort	☐
I have extreme pain or discomfort	☐
ANXIETY / DEPRESSION	
I am not anxious or depressed	☐
I am slightly anxious or depressed	☐
I am moderately anxious or depressed	☐
I am severely anxious or depressed	☐
I am extremely anxious or depressed	☐

- We would like to know how good or bad your health is TODAY.
- This scale is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Mark an X on the scale to indicate how your health is TODAY.
- Now, please write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐