Official Title of Study: An Open-label Randomized Multinational Phase 3 Trial of Nivolumab Versus Docetaxel in Previously Treated Subjects With Advanced or Metastatic Non-small Cell Lung Cancer (CheckMate 078: CHECKpoint Pathway and nivoluMAb Clinical Trial Evaluation 078)

NCT Number: NCT02613507

Document Date: May 2, 2017

STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT

AN OPEN-LABEL RANDOMIZED MULTINATIONAL PHASE III TRIAL OF NIVOLUMAB VERSUS DOCETAXEL IN PREVIOUSLY TREATED SUBJECTS WITH ADVANCED OR METASTATIC NON-SMALL CELL LUNG CANCER (NSCLC)

PROTOCOL CA209-078

VERSION # 2.1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

[redacted]

1.1 Research Hypothesis

The treatment effect of nivolumab is consistent between the study population and subjects from global pivotal trials (CheckMate 057 and CheckMate 017) and nivolumab increases overall survival (OS) as compared with docetaxel, in subjects with advanced or metastatic NSCLC who have failed prior platinum-based doublet chemotherapy.

1.2 Schedule of Analyses

OS is the primary endpoint for this study.

One formal interim analysis for superiority of the OS is planned when at least 291 deaths on all randomized subjects (76% of total deaths required for the final OS analysis) have been observed The interim analysis of OS will be monitored by an independent Data Monitoring Committee (DMC) and is expected to occur approximately 24 months after study initiation. Details are specified in the DMC charter.

The final analysis for OS is planned to be performed when at least 382 deaths are observed (approx 37 months after study initiation). Additional details can be found in Section 7.5.6.

In addition to the OS interim and final analyses, a Time to Treatment Failure interim analysis (TTF IA) is added to support an early filing in China. TTF IA will occur when the first ~380 randomized subjects have been followed for at least 8 months (TTF population). The interim analysis of TTF will be reviewed by the DMC. ORR and safety (including death summary) will be analyzed descriptively for subjects from the TTF population.

2 STUDY DESCRIPTION

2.1 Study Design

This is an open-label, randomized, Phase 3 study in adult (≥ 18 years old) male and female subjects with advanced or metastatic NSCLC after failure of prior platinum-doublet chemotherapy. Approximately 500 subjects, will be randomized to nivolumab vs. docetaxel in a 2:1 ratio.

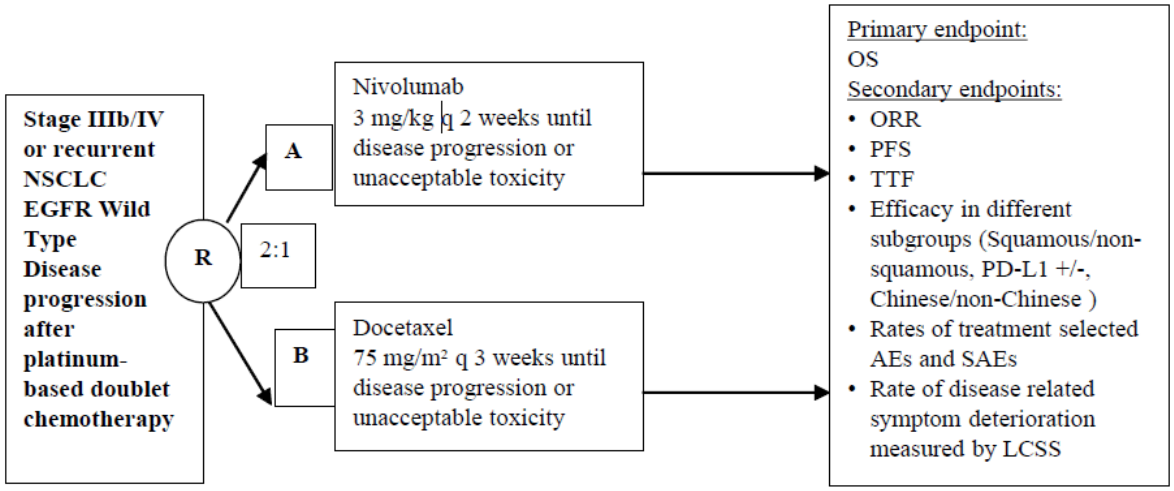
Subjects will undergo screening evaluations to determine eligibility within 28 days prior to randomization. (see Study Design and Duration schema in Figure 2.1-1 below). Randomization will be stratified and balanced according to the following factors: histology (squamous vs. non-squamous)/PD-L1 Status (positive vs. negative/unevaluable)/ ECOG Performance status (0 vs. 1). Approximately 100 PD-L1 unevaluable subjects will be randomized.

Treatment should be initiated within 3 business days of randomization. Nivolumab or docetaxel (depending on randomized treatment arm) will be administered as an IV infusion over 60 minutes on Treatment Day 1. A treatment cycle is defined as 2 weeks for nivolumab and 3 weeks for docetaxel.

Subjects will be evaluated for response according to the RECIST 1.1 criteria. The first on treatment radiographic assessment will be obtained in both treatment arms at Week 6 (± 7 days). The subsequent radiographic assessments will be conducted every 6 weeks (± 7 days) for the first 12 months (Week 48), then every 12 weeks (± 14 days) until disease progression (or until discontinuation of study therapy in patients receiving nivolumab beyond progression), lost to follow-up, withdrawal of study consent. Any subject who develops an objective tumor response per RECIST 1.1 (Complete Response or Partial Response) is required to undergo confirmatory scans at the next scheduled tumor assessment in the first year, and at least 4 weeks apart in the second and following year.

Subjects treated on the nivolumab arm will be allowed to continue study therapy after initial investigator-assessed RECIST 1.1-defined progression if they are assessed by the investigator to be deriving clinical benefit and tolerating study drug. Such subjects should discontinue study therapy when further progression is documented. After treatment discontinuation, subjects will have two follow-up visits for safety within the first 100 days from the last dose of study therapy. Beyond 100 days subjects will be followed for ongoing drug-related adverse events until resolved, return to baseline or deemed irreversible, or until loss of follow-up, withdrawal of study consent. All subjects will be followed for overall survival every 3 months until death, lost to follow-up or withdrawal of study consent.

Figure 2.1-1: Study Design



2.2 Treatment Assignment

Subjects are enrolled using the Interactive Voice Response System (IVRS) to obtain a subject ID. Subjects who have signed informed consent and met all eligibility criteria will be ready to be randomized through the IVRS, upon confirmation of receipt of required tissue sample by the central lab. The following information is required for subject randomization:

- Subject number
- Date of birth
- Gender at birth
- Date of informed consent
- Histology (squamous vs. non-squamous):
  - Subjects with mixed histology should be classified according to the predominant histology.
  - Subjects with adenosquamous histology should be classified as non-squamous histology
- ECOG Performance status (0 vs. 1)
- PD-L1 Status

The IVRS will randomly assign the subject in a 2:1 ratio to either Arm A (nivolumab) or Arm B (docetaxel), stratified by the following factors:

- Histology (squamous vs. non-squamous)
- PD-L1 Status (positive vs. negative/ unevaluable)
- ECOG Performance status (0 vs. 1)

The randomization will be carried out via permuted blocks within each stratum.

2.3 Blinding and Unblinding

Not applicable. This is an open-label study.

2.4 Protocol Amendments

This SAP incorporates following protocol amendments.

Table 2.4-1: Protocol Amendments

[redacted]

2.5 Data Monitoring Committee

A data monitoring committee (DMC) will be instituted for this study. This committee will provide independent oversight of safety and efficacy considerations, study conduct and risk-benefit ratio. Following review, the DMC will recommend continuation, modification, or discontinuation of this study based on reported safety and efficacy data. A separate DMC charter describes the members and activities of this committee, as well as the proposed meeting schedule. Representatives of the Sponsor will serve only as coordinators of the committee, without having full member responsibilities or privileges. In addition, the Sponsor will independently review safety data in a blinded manner during the conduct of this trial to ensure that any safety issues are identified and addressed.

3 OBJECTIVES

3.1 Primary

- To demonstrate that the OS benefit of nivolumab in the study population with advanced or metastatic NSCLC after failure of prior platinum-based chemotherapy is consistent with benefit observed in global studies CheckMate 057 and CheckMate 017
- To compare the OS of nivolumab versus docetaxel in subjects with advanced or metastatic NSCLC after failure of prior platinum-based chemotherapy.

3.2 Secondary

- To compare the objective response rate (ORR) of nivolumab vs. docetaxel
- To compare progression-free survival (PFS) of nivolumab vs. docetaxel
- To compare the time to treatment failure (TTF) of nivolumab versus docetaxel
- To evaluate clinical efficacy (OS, ORR, and PFS) in different subgroups, including squamous and non-squamous NSCLC subgroups, PD-L1 positive and PD-L1 negative protein expression subgroups, and Chinese and non-Chinese subgroups.
- To evaluate rates of treatment related selected AEs and SAEs in nivolumab and docetaxel arms.
- To evaluate the proportion of subjects exhibiting disease-related symptom deterioration by 12 weeks and 24 weeks, as measured by LCSS, in nivolumab and docetaxel arms.

4 ENDPOINTS

4.1 Primary Endpoint

OS is defined as the time from randomization to the date of death. A subject who has not died will be censored at last known date alive.

Consistency in OS benefit is defined as maintaining 50% of the risk reduction of death from CheckMate 057 and CheckMate 017 after adjusting for the patient distribution and the use of anti-PD-1/PD-L1 agents in the docetaxel arm. Method is described in detail in Section 5: Sample Size and Power

4.2 Secondary Endpoints

4.2.1 Objective Response Rate

Objective response rate (ORR) is defined as the number of subjects whose best objective response (BOR) is a confirmed CR or confirmed PR, as determined by the investigator, divided by the number of randomized subjects. BOR is defined as the best response designation, recorded between the date of randomization and the date of objectively documented progression per RECIST v1.1 or the date of subsequent anticancer therapy (excluding on-treatment palliative radiotherapy of CNS, skin or bone non-target lesions), whichever occurs first. For subjects without documented progression or subsequent anticancer therapy, all available response designations will contribute to the BOR determination. For subjects who continue treatment beyond progression, the BOR will be determined based on response designations recorded up to the time of the initial RECIST 1.1-defined progression.

4.2.2 Progression Free Survival

Progression-free survival (PFS) is defined as the time from randomization to the date of the first documented tumor progression as determined by the investigator using RECIST 1.1 criteria or death due to any cause. Clinical deterioration in the absence of unequivocal evidence of progression (per RECIST 1.1) is not considered progression for purposes of determining PFS. Subjects who die without a reported prior progression will be considered to have progressed on the date of their death. Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment. Subjects who did not have any on-study tumor assessments and did not die will be censored on the date they were randomized. Subjects who received any subsequent anti-cancer therapy (including on-treatment palliative RT of non-target bone lesions, skin lesions or CNS lesions) without a prior reported progression will be censored at the last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy.

Table 4.2.2-1: Censoring Scheme for Primary Definition of PFS

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
No on study tumor assessments and no death	Randomization	Censored
New anticancer treatment started without a prior reported progression per RECIST 1.1 or death	Date of last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored
Progression per RECIST 1.1 documented at scheduled or unscheduled visit and no new anticancer treatment started before	Date of the first documented tumor progression	Progressed
Subject progression free (per RECIST 1.1) and no new anticancer treatment started	Date of last tumor assessment	Censored

Table 4.2.2-1: Censoring Scheme for Primary Definition of PFS

Situation	Date of Progression or Censoring	Outcome
Death without prior progression per RECIST 1.1 and no new anticancer treatment started	Date of death	Progressed

4.2.3 Time to Treatment Failure

Time to Treatment Failure (TTF) is defined as the minimum of the time from randomization to the following dates:

- disease progression date (RECIST or clinical)
- death date
- last dose date if subject discontinued from treatment for any reasons other than "maximum clinical benefit", "administrative reasons by sponsor".

TTF is considered as event at the randomization date for subjects who were randomized but not treated. Clinical progression date is considered for time to treatment failure only when treatment is discontinued due to clinical disease progression. For nivolumab subjects treated beyond RECIST 1.1 progression, the event will be at RECIST 1.1 progression date. TTF is censored at the last dose date for subjects who discontinued treatment (without RECIST 1.1 progression) due to maximum clinical benefit or administrative reason by sponsor. TTF is censored at the last dose date for subjects who continued on treatment without progression or death. TTF will be analyzed at the interim analysis of TTF only.

4.2.4 Disease related Symptom Deterioration Rate by Week 12 and Week 24

Disease-Related Symptom deterioration Rate by Week 12 and Week 24 is defined as the proportion of randomized subjects who had 10 points or more increase from baseline in Average Symptom Burden Index (ASBI) score at anytime between randomization and week 12 /week 24 respectively.

The LCSS is a measure of disease-related symptoms and quality of life suited to use in patients suffering from lung cancer. It includes six items measuring loss of appetite, fatigue, coughing, shortness of breath, hemoptysis, and pain. Three additional items measure overall symptom burden, disease-related functional limitations, and quality of life. The questionnaire uses a 24-hour recall period, and responses for each item are captured using a 100mm visual analog scale (VAS). Scores for individual items ranging from 0 (no symptomatology or highest quality of life) to 100 (worst symptomatology or quality of life) are derived by dividing the length of the line drawn from the lowest possible response to the patient's response by the length of the VAS and multiplying the resulting quotient by 100. An average symptom burden index (ASBI) score can be derived as the average of scores for the six symptom-related items with a clinically meaningful change in

ASBI score being defined as 10 points. Accordingly, a meaningful deterioration in symptoms as measured by the ASBI is reflected in a mean post-baseline score change ≥ 10 points.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5 SAMPLE SIZE AND POWER

The sample size is calculated in order to compare the OS between subjects randomized to receive nivolumab and subjects randomized to receive docetaxel. Approximately 500 subjects will be randomized to the nivolumab and docetaxel arms in a 2:1 ratio. At least 382 deaths will be required for the final analysis of OS. A formal OS interim analysis will be conducted when at least 291 deaths (76% information fraction) have been observed.

Simulations were run in R to calculate the power of the (weighted) log-rank test.

A mixed population was generated as follows: 500 subjects were randomly assigned to 3 cohorts with 40% squamous (SQ) subjects, 30% non-squamous (NSQ) PD-L1 positive subjects and 30% non-squamous (NSQ) PD-L1 negative subjects.

Heterogeneity of treatment effects across histology (SQ vs. NSQ) and PD-L1 status subgroups was considered with the assumptions as described below:

- In the docetaxel treatment group, mOS was assumed to be 10 months for SQ and 12 months for NSQ, as per Asian data.[4,5,6,7,8,9] Exponential distributions of OS times were assumed in both histologies.
- For SQ subjects from nivolumab treatment group, a hazard ratio of 0.6 (mOS nivolumab vs. docetaxel: 16.7 vs. 10 months) and exponential distribution of OS times were assumed.

Assumption on treatment effect was based on results from the Phase 3 study CA209017 (HR: 0.59; 95% Condidence Interval (CI): 0.44 - 0.79).

- For NSQ subjects from nivolumab treatment group, non proportional hazard models was assumed based on results from the Phase 3 study CA209057. Piecewise exponential models were used to model survival function. For the NSQ PD-L1 positive subjects, a 4-month delay effect (exponential distribution with mOS of 12 months) followed by an exponential distribution with mOS of 24 months was used in the model, providing an overall mOS of 20 months. For the NSQ PD-L1 negative subjects, the piecewise exponential model was as followed: for the first 2 months, survival function was assumed to follow an exponential distribution with mOS of 5 months; after Month 2, an exponential distribution with mOS of OS 18 months, providing an overall mOS of 12.8 months.

Overall survival (OS) in all comers was compared between nivolumab and docetaxel, using a two-sided log-rank test with a significance level of 5%. Fleming-Harrington (FH) weighted log-rank test was also used as as it is known to be more efficient for testing survival difference when a delay effect is present. The power of both the log-rank and FH weighted log-rank tests, using G (rho = 0, gamma = 1) weights in the terminology of Fleming and Harrington was assessed. To control the overall Type I error rate under a two-sided 5%, significance levels of 0.020 and 0.044 were used at the interim and final analyses to decide if the study could be stopped for efficacy. These significance levels were calculated using the Lan-DeMets error spending function by East (version 6). Hazard ratios (HR) were estimated using Cox proportional hazards model with treatment arm as the only covariate in the model.

At both IA OS and FA OS, a consistency check will be performed first followed by superiority test (hierarchical testing). Consistency was defined if the observed HR for OS maintains at least 50% of the risk reduction of death in the global studies. The probability that the observe HR is not greater than the consistency HR threshold was calculated at both the interim and final analyses i.e. the probability that the simulated HR ≤ the consistency HR among all simulated trials.

The impact of crossover effect was considered. Power and hazard ratios were calculated assuming 0%, 5%, 10%, 15% and 20% crossover rates in the control arm. For subjects in the control arm who were identified as crossover, their survival time were simulated based on the OS distribution of the treatment arm according to their histology and PD-L1 status

All results were generated by 5000 simulations. Accrual information used in the simulations had the same pattern as the actual data at time of the protocol amendment (500 subjects were accrued in 11 months).Simulation results are displayed in Table 5-1

In absence of cross over, IA OS (291 deaths) and FA OS (382 deaths) occurred approximately at 24 months (13 months of minimum follow-up) and 37 months (26 months of minimum follow-up) respectively. Using the log-rank test, cumulative power was 73% and 96% and average HR was 0.70 and 0.67 at the interim and final analyses, respectively. Power reached 86% and 98% accordingly with the weighted log-rank test. The Probability of Technical Success (PTS) for consistency check at the interim analysis was 95% (ie. the probability that the observed HR ≤ 0.850). At the final analysis, probability of technical success for consistency check was 98% (ie.

probability that the observed HR ≤ 0.835 was 98%). PTS of the log-rank test decreases when cross-over is present. PTS of the weighted log-rank test decreases as well but less than the conventional log-rank test. PTS for consistency check remains above 90% for crossover rates up to 20%.

Table 5-1: Simulations results

	No Crossover	5% Crossover	10% Crossover	15% Crossover	20% Crossover
Events at IA/FA	291/382	291/382	291/382	291/382	291/382
Timing for IA/FA (months)	24/37	24/38	24/38	24/38	24/38
Expected HRs at IA/FA	0.70/0.67	0.71/0.69	0.73/0.71	0.74/0.72	0.75/0.73
Consistency HR thresholds at IA/FA (maintaining 50% of risk reduction)	0.850/0.835	0.855/0.845	0.865/0.855	0.870/0.860	0.875/0.865
PTS for log-rank test at IA/FA	73%/96%	69%/94%	63%/90%	59%/86%	53%/82%
PTS for weighted log-rank test at IA/FA	86%/98%	82%/97%	77%/94%	72%/91%	67%/88%
PTS for consistency check	95%/98%	93%/98%	93%/97%	91%/96%	90%/94%

After the first ~380 randomized subjects have been followed for at least 8 months (approx. 16 months after study initiation), TTF is to be compared across treatment groups using a weighted log-rank test[10, 11]. With one-sided type 1 error of 0.025, the power for the weighted log-rank test of TTF was ~95% per simulations using data observed in global pivotal studies (CheckMate 057 and 017), and , adjusted for the proportion of SQ and NSQ in this study. Expected HR for TTF was 0.7.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

See Core Safety SAP.

6.2 Treatment Regimens

The treatment group "**as randomized"** will be retrieved from the IVRS system

- Arm A: Experimental arm (monotherapy) nivolumab
- Arm B: Control arm docetaxel

The treatment group "**as treated**" will be the same as the arm randomized by IVRS. However, if a subject received the incorrect drug for **the entire period** of treatment, the subject's treatment group will be defined as the incorrect drug the subject actually received.

6.3 Populations for Analyses

- **All enrolled subjects**: All subjects who signed an informed consent form and were registered into the IVRS. Analyses of the patients enrolled into the study but not randomized and the reason for not being randomized will be performed on the data set of all enrolled subjects.
- **All randomized subjects**: All subjects who were randomized to any treatment group in the study. This is the primary dataset for analyses of demography, protocol deviations, baseline characteristics, efficacy, outcome research.
- **TTF population:** All randomized subjects with at least 8 months of follow-up at clinical database cutoff date of TTF interim analysis (~380 subjects). This dataset will be used for all ITT analyses at the TTF IA.
- **All treated subjects**: All subjects who received at least one dose of nivolumab or docetaxel. This is the primary dataset for dosing and safety.
- **Treated subjects among TTF population**: This is the primary dataset safety and exposure analyses on the treated subjects among TTF population.
- **All responders:** randomized subjects with CR or PR
- **Response evaluable subjects**: randomized subjects whose change in the sum of diameters of target lesions was assessed (ie.: target lesion measurements were made at baseline and at at least one on-study tumor assessment).
- **PK subjects**: All subjects with available serum time-concentration data from randomized subjects dosed with nivolumab.
- **Immunogenicity subjects**: See Core Safety SAP.
- **Biomarker subjects**: All randomized subjects who had available biomarker data

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, the bulleted titles in the following subsections describe tabulations of discrete variables, by the frequency and proportion of subjects falling into each category, grouped by treatment (with total). Percentages given in these tables will be rounded and, therefore, may not always sum to 100%. Continuous variables will be summarized by treatment group (with total) using the mean, standard deviation, median, minimum and maximum values.

Time to event distribution (i.e. time to treatment failure, progression free survival, overall survival and duration of response) will be estimated using Kaplan Meier techniques.

Median survival time along with 95% CI will be constructed based on a log-log transformed CI for the survivor function S(t)[12,13]. Rates at fixed timepoints (e.g. OS at 6 months) will be derived from the Kaplan Meier estimate and corresponding confidence interval will be derived based on Greenwood formula[14] for variance derivation and on log-log transformation applied on the survivor function S(t)[15].

Unless otherwise specified, the stratified log-rank test will be performed to test the comparison between time to event distributions. Stratification factors will be prior use of histology (squamous

vs. non-squamous)/ PD-L1 Status (positive vs. negative/unevaluable)/ ECOG Performance status (0 vs. 1), as entered into the IVRS.

Unless otherwise specified, the stratified hazard ratio between 2 treatment groups along with CI will be obtained by fitting a stratified Cox model with the treatment group variable as unique covariate. Stratification factors will be same as above.

The difference in rates between the two treatment arms along with their two-sided 95% CI will be estimated using the following Cochran-Mantel-Haenszel (CMH) method of weighting[16], adjusting for the stratification factors:

$$\hat{\theta} = \frac{\sum_i w_i \hat{\theta}_i}{\sum_i w_i} \sim N\left[\theta, \frac{\sum_i w_i^2 \left[\frac{p_{ix}(1-p_{ix})}{n_{ix}-1} + \frac{p_{iy}(1-p_{iy})}{n_{iy}-1}\right]}{\left(\sum_i w_i\right)^2}\right]$$

where $\hat{\theta} = p_{ix} - p_{iy}$ is the rate difference of the ith stratum, $w_i = \frac{n_{ix} n_{iy}}{n_{ix} + n_{iy}}$, and n_{ix} and n_{iy} are the number of subjects randomized to treatments x and y, respectively, in the ith stratum.

Stratification factors will be the same as above. Associated odds-ratio will be derived.

The p-values from sensitivity analyses for efficacy endpoints are for descriptive purpose only and not adjusted for multiplicity.

7.2 Study Conduct

Unless otherwise specified, study conduct analyses at the TTF analysis timepoint will be performed using the TTF population. At the formal OS analysis timepoints, analyses will be conducted using the all randomized population.

7.2.1 Accrual

The accrual pattern will be summarized per country, investigational site and per month for all enrolled subjects. Randomization date, first dosing date, country, investigational site will be presented in a by subject listing of accrual.

7.2.2 Relevant Protocol Deviations

The relevant Protocol Deviations will be summarized, by treatment group and overall. The following programmable deviations from inclusion and exclusion criteria will be considered as relevant protocol deviations. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and non-programmable) eligibility and on-treatment protocol deviations will be reported through ClinSIGHT listings.

At Entrance:

- Subjects who received first-line treatment but not with platinum doublet chemotherapy.
- Subjects who didn't receive first-line treatment.
- Subjects with known EGFR mutation or ALK rearrangement.
- Subjects who received prior treatment with docetaxel.
- Subjects without measurable disease at baseline.
- Subject with baseline ECOG PS > 1.

On-study:

- Subjects receiving concurrent anti-cancer therapy (defined as chemotherapy, hormonal immunotherapy, radiation therapy, standard or investigational agents for treatment of NSCLC).
- Subjects treated differently as randomized (subjects who received the wrong treatment, excluding the never treated).

A subject listing will also be produced.

7.3 Study Population

Unless otherwise specified, study population analyses at the formal TTF analysis timepoint will be performed on the TTF population . At the formal OS analysis timepoints, analyses will be performed on the all randomized population.

7.3.1 Subject Disposition

The total number of subjects enrolled (randomized or not randomized) will be presented along with the reason for not being randomized.

Number of subjects randomized will be presented along with reason for not being treated.

Days from Randomization to first dose date (≤ 3, 4 - 5, 6-7, 8-14, 15 - 21, > 21 days) will be tabulated for all treated subjects.

Number of treated subjects by study treatment discontinuation status will be tabulated along with reason for treatment discontinuation.

The current status of follow-up for overall survival will be summarized for all randomized subjects. Subjects count by for time from date of last contact to last patient last visit will be presented using the following categories: 0, 1 - 30, 31 - 60, 61 - 90, 91 - 120, 121 - 150, > 150 days. Subjects who have died will categorized as having an event at day 0. For subjects with last known date alive after data cut-off date, they will have zero value for currentness of follow-up as well.

7.3.2 Demographics and Other Baseline Characteristics

The following baseline characteristics will be summarized by treatment arm as randomized. All baseline presentations identify subjects with missing measurements. Listings will also be provided. Source is CRF.

- Age (descriptive statistics).
- Age category (< 65, ≥ 65 - < 75, ≥ 75, ≥ 65).
- Gender (male/female)
- Race (White/Black/Asian/Other)
- Region (Asia vs. Europe).
- Country (China (split by China Mainland/Hong Kong), Singapore, Russia)

- Baseline ECOG Performance Status
- Weight (Kg)..
- Smoking Status (current/former smoker, never smoker, unknown)
- Cell Type (Squamous cell carcinoma, Adenocarcinoma, Large cell carcinoma, Other-specify)
- Time from Initial Disease Diagnosis to Randomization (< 1 year, 1-< 2 year, 2 - < 3 year, 3 - < 4 year, 4 - < 5 year, ≥ 5 year).
- Disease stage at study entry (stage IIIB, stage IV, recurrent)
- EGFR mutation status, ALK translocation status, KRAS mutation status.
- All lesions (Investigator Tumor Assessments at Baseline): sites of disease, number of disease sites per subject.
- Target Lesions (Investigator Tumor Assessments at Baseline): Presence of target lesions, site of target lesion, sum of diameters of target lesions.
- CNS metastases (yes/no).

7.3.3 Medical History

General medical history will be listed by subject and pretreatment events will be tabulated.

7.3.4 Prior Therapy Agent

Prior anti-cancer therapy will be summarized as follows:

- Number of lines of prior systemic cancer therapy received (1, 2, > 2)
- Number of Subjects who received prior maintenance therapy
- Prior systemic therapy regimen setting (number of subjects who received adjuvant therapy, neo-adjuvant therapy, number of subjects who received regimen with metastatic setting)
- Prior systemic cancer therapy summary (by agent)
- Best response to most recent prior systemic therapy regimen (CR/PR, SD, PD, Unknown or Not reported)
- Time from completion of most recent prior systemic therapy regimen to randomization (< 3 months, 3 - 6 months, > 6 months)
- Time from completion of prior adjuvant/neo adjuvant to randomization (< 6 months, ≥ 6 months)
- Prior surgery related to cancer (yes, no, not reported)
- Prior radiotherapy (yes, no, not reported)

Prior systemic therapy will be classified by therapeutic class and generic name.

Other Prior therapy:

- Prior/current non-study medication classified by anatomic and therapeutic classes.

Agents and medication will be reported using the generic name. A listing by subject will also be provided.

7.3.5 Baseline Examinations

Subjects with abnormal baseline physical examination will be tabulated by examination criteria and by treatment group.

7.3.6 Discrepancies between IVRS and CRF Stratification Factors

Summary tables (cross-tabulations) by treatment group for the following stratification factor will be provided to show any discrepancies between what was reported through IVRS vs. CRF data (baseline).

- ECOG Performance Status: 0 vs. 1
- Histology: Squamous , Non-Squamous

PD-L1status (Positive vs. Negative/unevaluable) and PD-L1 expression will be reported as entered into the IVRS.

Frequency of PD-L1 Status at baseline (Source: IVRS) will be summarized using 3 pre-specified PD-L1 expression cutoff values (1%, 5% and 10%).

- PD-L1 Positive: subjects with baseline PD-L1 expression ≥ x% cutoff value
- PD-L1 Negative: subjects with baseline PD-L1 expression < x% cutoff value and ≥ 0%
- PD-L1 Unevaluable: subjects with no quantifiable PD-L1 expression at baseline.

7.4 Extent of Exposure

Extent of exposure analyses will be performed by treatment group “as treated”. The primary population at theTTF analysis will be the Treated subjects among TTF population. The population at the formal OS analyses will be the all treated population.

7.4.1 Administration of Study Therapy

The following parameters will be summarized (descriptive statistics) by treatment group:

- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%.
- Number of doses received (summary statistics).
- Cumulative dose

Duration of treatment will be presented by treatment group using a Kaplan-Meier curve whereby the last dose date will be the event date for those subjects who are off study therapy. Median duration of treatment and associated 95% CI will be provided. Subjects who are still on study therapy will be censored on their last dose date.

A by-subject listing of dosing of study medication (record of study medication, infusion details, dose change) and a listing of batch number will be also provided.

Below table summarizes the key parameters used to calculate dosing data.

Table 7.4.1-1: Administration of Study Therapy: Definition of Parameters

	Nivolumab	Docetaxel
Dosing schedule per protocol	3mg/kg every 2 weeks	75mg/m² every 3 weeks
Dose	Dose (mg/kg) is defined as total dose administered (mg)/most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.	Dose (mg/m²) is is defined as total dose administered (mg)/most recent BSA. Dose administered in mg at each dosing date and BSA (computed using recent weight and baseline height) are collected on the CRF.
Cumulative Dose	Cum dose (mg/kg) is sum of the doses (mg/kg) administered to a subject during the treatment period.	Cum dose (mg/m²) is sum of the doses (mg/m²) administered to a subject during the treatment period.
Relative dose intensity (%)	Cum dose (mg/kg) / [(Last dose date - Start dose date + 14) × 3 / 14] × 100]	Cum dose (mg/m²) / [(Last dose date - Start dose date + 21) × 75 / 21] × 100]
Duration of treatment	Last dose date - Start dose date + 1	Last dose date - Start dose date + 1

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delays

A dose will be considered as actually delayed if the delay is exceeding 3 days (i.e., greater than or equal to 4 days from scheduled dosing date) for both nivolumab and docetaxel. It is defined as (duration of previous cycle in days -14) for nivolumab (or -21 for docetaxel). Dose delays will be divided into following categories: on-time, 4 - 7 days, 8 - 14 days, 15 - 42, > 42 days. Reason for dose delay will be retrieved from CRF dosing pages.

Each nivolumab or docetaxel infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment arm:

- Number of dose delayed per subject, Length of Delay and Reason for Dose Delay.
- Number of subject with at least one dose infusion interrupted along with reason for the interruptions and number of infusions interrupted per subject.

7.4.2.2 Dose Reductions

There will be no dose escalations or reductions of nivolumab allowed. Subjects may be dosed no less than 12 days from the previous dose.

Each nivolumab or docetaxel infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages.

Dose of Docetaxel may be modified for toxicity. Subjects may be dosed no less than 19 days from the previous dose.

Dose levels of docetaxel are defined in the protocol as follows:

- Dose level -1: 55mg/m².
- Dose level -2: 37.5mg/m².

For any cycle (excluding Cycle 1), it will be defined as a dose reduction if the observed dose level (based on calculated administered dose) is below the dose level of the previously administered dose. Dose ranges for dose levels of docetaxel are defined in Table 7.4.2.2-1.

Table 7.4.2.2-1: Calculated Dose Ranges and Related Dose Levels of Docetaxel

Dose Range (mg/m^2)	Dose Level
≥ 65	Level 0
≥ 45 - < 65	Level -1
< 45	Level -2

The reason for dose reduction as reported by the investigator will be tabulated for all instances of dose reduction based on the Dose Change CRF page. A category 'Unknown' will be defined for all reductions with no reason reported by the investigator.

The following will be summarized by treatment group:

- Number of subjects with at least one infusion with IV rate reduced along with the reason of the rate reduction.

The following will be summarized for docetaxel subjects:

- Number and percentage of subjects with at least one dose reduction and reason of the dose reduction, number and percentage of subjects with a dose reduction to dose level -1, number and percentage of subjects with a dose reduction to dose level -2.

7.4.3 Concomitant Medications

Concomitant medications, defined as medications other than study medications which are taken at any time on-treatment (i.e. on or after the first day of study therapy and within 100 days following the last dose of study therapy), will be coded using the WHO Drug Dictionary.

The following summary tables will be provided:

- Concomitant medications (subjects with any concomitant medication, subjects by medication class and generic term).

A by-subject listing will accompany the table.

7.5 Efficacy

Efficacy analyses at the TTF analysis timepoint will be performed on the TTF population. At the formal OS analysis timepoints, analyses will be performed on the all randomized population.

7.5.1 Overall Survival

7.5.1.1 Primary Analysis

OS is the primary endpoint of this study.

At both the interim OS analysis and final OS analysis, a 2-step hierarchical testing will be performed. First, a check for consistency in HR for OS will be performed. This is to mitigate loss of power due to confounding effects of cross-over (subjects receiving anti PD-1/PD-L1 agents in the comparator arm).

If consistency in OS with global data is demonstrated, superiority for OS will be tested. The distribution of OS will be compared in two randomized arms via a two-sided, using a two-sided α (adjusted for the interim) weighted log-rank test stratified by histology (squamous vs. non-squamous)/ PD-L1 Status (positive vs. negative/unevaluable)/ ECOG Performance status (0 vs. 1)., as entered into the IVRS. The weighted log-rank test will use G (rho = 0, gamma = 1) weights, in the terminology of Fleming and Harrington[10,11].

The unweighted hazard ratio (HR) and the corresponding 100(1-α)% CI (adjusted for the interim) will be estimated in a stratified Cox proportional hazards model using randomized arm as a single covariate with the same stratification factors mentioned above.

The OS curves for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. Two-sided, 95% confidence intervals for median OS will be constructed based on a log-log transformed CI for the survivor function S(t) [12,13].

Survival rates at 6, 12, 18, 24, 36, 48 months and at 5 year will also be estimated using KM estimates on the OS curve for each randomized arm. Minimum follow-up must be longer than timepoint to generate the rate. Associated two-sided 95% CIs will be calculated using the Greenwood's formula[14] for variance derivation and on log-log transformation applied on the survivor function S(t)[15].

The status of subjects who are censored in the OS Kaplan-Meier analysis will be tabulated for each treatment group using following categories:

- on-study (on-treatment and not progressed, on-treatment progressed, in follow-up);
- off-study: (lost to follow-up, withdraw consent, etc.).

To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. A two-sided Wald Chi-square p-value of less than 0.1 may indicate a potential nonconstant treatment effect. In that case, additional exploratory analyses may be performed.

7.5.1.2 OS Sensitivity Analyses

The following OS sensitivity analyses will be performed.

- OS will be compared between treatment groups using a two-sided α (adjusted for the interim) stratified regular log-rank test.
- OS will be compared between treatment groups using a two-sided α (adjusted for the interim) unstratified regular log-rank test.
- OS will be compared between treatment groups using a two-sided α (adjusted for the interim) stratified regular log-rank test, and using the strata as determined at baseline (CRF source). This analysis will be performed only if at least one stratification variable at IVRS and at baseline disagrees for at least 10% of the randomized subjects.
- OS will be compared between treatment groups using a two-sided α (adjusted for the interim) stratified regular log-rank test in the All Treated Subjects population, using arm as randomized. This analysis will be performed only if the proportion of randomized but never treated subjects exceeds 10% in any arm.
- OS will be compared between treatment groups using a two-sided α (adjusted for the interim) stratified regular log-rank test, but censoring in chemo arm, subjects at the subsequent use of anti PD(L)1.

Estimate of the hazard ratio, its two sided 100(1-α)% CI (adjusted for the interim) and p-value will be presented.

- OS distribution will be compared bewteen treatment groups using stratified log-rank test limited to the first 291 deaths. Kaplan-Meier curves will be produced and stratified HR along with the two sided 100(1-α)% CI (adjusted for 291 deaths) will be computed.

7.5.1.3 Consistency of Treatment Effect on OS in Subsets

To assess consistency of treatment effects in different subsets, a "forest" plot of the OS unstratified hazard ratio (and 95% CI) will be produced for the following subgroups.

- Age categorization (< 65, ≥ 65 - < 75, ≥ 75, ≥ 65).
- Gender (male, female).
- Race (White, Asian, Other).
- Chinese / Non-Chinese., based on country (Chinese = from China Mainland or Hong-Kong)
- Baseline ECOG PS (0 vs. ≥ 1).
- Smoking status (CURRENT/FORMER vs. NEVER SMOKED vs. unknown).
- Histology (SQ vs. NSQ) - per CRF
- PD-L1 status at baseline using 1% cutoff
- Disease Stage at study entry (stage IIIB vs. stage IV/recurrent)
- Time from initial diagnosis to randomization (< 1 year vs. other)).
- Time from completion of most recent prior systemic therapy regimen to randomization: < 3 months, 3-6 months, > 6 months, unknown.
- Prior maintenance therapy (yes vs. no).
- Best response to most recent prior regimen (responders (CR/PR) vs. non responders (SD vs. PD) vs. UTD/NA)
- CNS metastases (yes/no).

If subset category has less than 10 subjects per treatment group, HR will not be computed/displayed. Number of events and median OS along with 95% CI will be displayed for each treatment group.

7.5.1.4 Multivariate Analysis

A multivariate stratified (by histology (squamous vs. non-squamous), PD-L1 Status (positive vs. negative/unevaluable), ECOG Performance status (0 vs. 1) as entered in the IVRS) Cox model will be fitted to assess the treatment effect on OS when adjusted for potential prognostic factors. The following potential prognostic factors will be included in the model.

- Time from initial diagnosis to randomization (< 1 year (yes vs. other))
- Age categorization (< 65, ≥ 65)
- Gender (Male vs. Female)
- Smoking status (yes vs. no or unknown)
- Disease Stage at study entry (stage IIIB vs. stage IV/recurrent)

HR and 95% CI will be provided for treatment variable and all covariates. Descriptive p-values will be provided.

7.5.1.5 Subject Follow-up for OS

The minimum follow-up will be reported. The minimum follow-up is defined as the time interval between the last patient's randomization date and the clinical cutoff date.

The extent of follow-up defined as the time between randomization date and last known date alive (for subjects who are alive) or death date (for subjects who died) will be summarized descriptively (median, min, max) for all subjects randomized.

The currentness of follow-up, defined as the time between last OS contact (i.e., last known date alive or death date) and data cut-off date, will be summarized by treatment group. Subjects who died before data cut-off date will automatically have zero value for currentness of follow-up. For subjects with last known date alive after data cut-off date, they will have zero value for currentness of follow-up as well. The currentness of follow-up will be categorized into the following categories: 0 days, 1-3 months, 3-6 months, 6-9 months, 9-12 months and ≥ 12 months.

7.5.1.6 Follow-up Therapy

Subsequent therapies will be summarized and listed.

- Subsequent Therapy
  - Immunotherapy (anti-PD1, anti PDL1, anti-CTLA4 and other Immunotherapy, by drug name)
  - Targeted Therapy (ALK/EGFR Tyrosine kinase inhibitors, VEGFR Inhibitors, other, by drug name)
  - Chemotherapy by drug name
  - Other investigational agent by drug name
  - Surgery
  - Radiotherapy
  - Any combination of the above
- By Subject Listing of Subsequent Therapy

7.5.1.7 Analysis of Survival by Tumor Response

Survival by response category will be analyzed by arm using the landmark method[17]. Subjects still on study at the landmark time will be separated into two response categories according to whether they have responded before that time. This will assess whether survival from the landmark depends on the subject's response status at the landmark. Subjects who go off protocol (e.g. subjects who die) before the time of landmark will be excluded from the analysis.

The survival curves from Week 9, Month 4, Month 6, Month 8, Month 12, by response status, for each randomized arm will be produced using the KM product-limit method. Two sided, 95% confidence intervals for median OS will be constructed based on a log-log transformed CI for the survivor function S(t).

7.5.1.8 OS Analysis at time of TTF analysis

OS won't be analyzed at time of interim analysis of TTF.

7.5.2 Objective Response Rate

7.5.2.1 Primary Analysis of ORR

BOR will be summarized by response category for each treatment group. ORR will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method[18].

An estimate of the difference in ORRs and corresponding 95% CI will be calculated using CMH methodology and adjusted by the same stratification factors as for primary analysis of OS.

In addition, the stratified (source: IVRS) odds ratios (Mantel-Haenszel estimator) between the treatments will be provided along with the 95% CI. The difference will be tested via the Cochran Mantel-Haenszel (CMH) test using a two-sided, 5% α level.

To assess consistency of treatment effect on ORR in different subsets, ORR will be computed across the same subsets as defined in the OS analysis, including squamous and non-squamous subgroups, PD-L1 positive and PD-L1 negative protein expression subgroups, and Chinese and non-Chinese subgroups (see Section 7.5.1.3).

7.5.2.2 Sensitivity Analyses of ORR

If one stratification variable at IVRS and at baseline (CRF) disagrees for at least 10% of the randomized subjects, similar analysis of ORR as primary analysis will be performed using the strata as determined at baseline.

7.5.2.3 Duration of Objective Response

Duration of response in each treatment group will be estimated using KM product-limit method for subjects who achieve PR or CR. Duration of stable disease will also be estimated in each treatment groups using KM product-limit method for subjects with SD as best response. Median values along with two-sided 95% CI will be calculated. Summary statistics will be computed constructed based on a log-log transformed CI for the survivor function S(t).

7.5.2.4 Time to Objective Response

Summary statistics of time to objective response will be provided for each treatment group for subjects who achieve PR or CR.

To assess further tumor response kinetics, time to response will be analyzed using the KM methology, for all randomized subjects. Kaplan-Meier curve will represent the cumulative rate of response over time. For the non-responders, time to response will be censored at the maximum time of response + 1 day of all subjects in their respective treatment group. Cumulative Response Rates will be tabulated for Week 9, Month 4, 6, 8, and 12, and overall Response Rate will be provided for each treatment group.

7.5.2.5 ORR Analysis at time of TTF analysis

At the TTF analysis, BOR will be summarized by response category for each treatment group. ORR will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method. Population for analysis will be the TTF population. Additionally analyses will be repeated using the Chinese subjects among the TTF population, and ORR will also be assessed by histology and, by PDL1 status at baseline (using cutoff at 1%, 5% and 10%) . TTR and DOR will be evaluated for responders (i.e. subjects with confirmed CR or PR) of the TTF population.

7.5.3 Progression Free Survival

7.5.3.1 Primary Analysis of PFS

PFS for each treatment arm will be estimated using Kaplan-Meier product limit method and graphically displayed. A two-sided 95% CI for median duration will be constructed based on a log-log transformed CI for the survivor function S(t).

The comparison of PFS distribution will be performed via a stratified log-rank test at two-sided, 5% level. In addition, the stratified hazard ratios between treatment groups will be provided along with the 95% CI.

PFS rates at 6, 12, 18, 24, 36, 48 months and at 5 year will also be estimated using KM estimates on the PFS curve for each randomized arm. Minimum follow-up must be longer than or equal to timepoint to generate the rate. Associated two-sided 95% CIs will be calculated using the Greenwood's formula[14].

The source of progression (death vs. progression) will be summarized by treatment group.

The status of subjects who are censored in the PFS Kaplan-Meier analysis will be tabulated for each treatment group using following categories:

- Never treated
- On-study (on treatment, in follow-up)
- Off-study: (lost to follow-up, withdrew consent, other).
- Received subsequent anti-cancer therapy
- No baseline tumor assessment

To assess consistency of treatment effect on PFS in different subsets, a "forest" plot of the PFS unstratified hazard ratio (and 95% CI) will be produced for the same variable as in the OS analysis.

7.5.3.2 Sensitivity Analyses of PFS

Sensitivity analyses of PFS will also be performed using the following modification of PFS primary definition.

- PFS will be compared between treatment groups *using the strata as determined at baseline* (CRF source). This analysis will be performed only if at least one stratification variable at IVRS and at baseline disagrees for at least 10% of the randomized subjects.

- *PFS accounting for assessment after subsequent therapy* subjects will be defined similarly to the primary definition except that events (progression or death) and tumor assessments that occurred on or after subsequent anticancer therapy will be taken into account.
- *PFS accounting for missing tumor assessment prior to PFS event (progression or death)*. This analysis will be performed only if at least 20% of events have missing prior tumor assessment. It will apply the following restriction to the primary definition: If the elapsed time between the PFS event and the last on-study assessment immediately prior to the event (or randomization date if no on-study scan) is two or more missed visits (more than 12 weeks + 10 days), the subject will be censored at his last tumor assessment prior to the PFS event (or randomization date if no on-study scan).
- PFS *accounting for assessment after on-treatment palliative radiotherapy on skin, bone or CNS non-target lesions*. This analysis is similar to the primary analysis except that no censoring will occur for on-treatment palliative radiotherapy.

7.5.3.3 PFS Analysis at time of TTF analysis

PFS won't be analyzed at time of interim analysis of TTF.

7.5.4 Hierarchy for Key Secondary Efficacy Endpoints

In order to preserve an experimental-wise type I error rate at 5%, a pre-planned hierarchy for key secondary ORR and PFS endpoints will be applied when interpreting the statistical significance of treatment comparisons. The hierarchical ordering of the key secondary endpoints is as follows:

1) Objective Response Rate
2) Progression-Free Survival

The statistical testing will be carried out using the following sequential procedure:

- If superiority of OS is demonstrated at interim or final OS analysis or if consistency of treatment effect in OS with global data is demonstrated at OS final analysis , the secondary endpoint with the highest ranking in the hierarchy will be tested (ORR in this case). If the p-value of ORR is statistically significant at 5% level, the second highest ranking endpoint in the hierarchy will be tested (PFS in this case) and the p-value of PFS will be provided. If the p-value of ORR is not statistically significant at 5% level, then no further statistical testing regarding the other secondary endpoint (i.e., PFS) will be conducted. Estimates (medians and HR) and their 95% CI will be provided for PFS regardless of the outcome of ORR testing.
- If not, then no further statistical testing regarding the secondary endpoints will be conducted. However estimates along with their 95% CI will be provided for those (i.e. medians and HR, rates and odds ratio).

7.5.5 Other Efficacy Analyses

The following subject-level graphics will also be provided by treatment group as randomized:

- For the responders only, time courses of the following events of interest will be graphically displayed: tumor response, tumor progression, last dose received, and death.

- For response evaluable subjects, a waterfall plot showing the best reduction in target lesion will be produced.

7.5.6 TTF Interim Analysis

A Time to Treatment Failure (TTF) interim analysis is planned after the first ~380 randomized subjects have been followed for at least 8 months (TTF population). This is projected to occur approx. 16 months after study initiation.

The DMC will review the safety and efficacy data from the TTF interim analysis. There is no alpha spending nor penalty for TTF analysis: this interim analysis is considered as a bridging strategy. If TTF is statistically significant at 1-sided 0.025 level, study conduct, study population, exposure, safety (including death summary) and ORR will be described (no formal comparison) using the TTF population (see section 7.5.2.5) .

Regardless of TTF is positive or negative, study will continue until the planned OS interim/final analysis.

The distribution of TTF will be compared in the two randomized arms via a 1-sided, weighted log-rank test (unstratified). The one-sided weighted log-rank p-value will be reported using G (rho = 0, gamma = 1) weights, in the terminology of Fleming and Harrington[10],[11]. The FH method (with rho=0 and gamma=1) leads to a loss of power if treatment effect is not delayed but higher power starting with approximately a 2 month delay in treatment effect and increases as the delay increases. Distribution of TTF will also be tested using regular (unweighted) log-rank test.

The hazard ratio (HR) and the corresponding 2-sided 95% CI upper bound will be estimated in an unstratified cox proportional hazards model using randomized arm as a single covariate. The TTF curves for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. Two-sided, 95% confidence intervals for median TTF will be constructed based on a log-log transformed CI for the survivor function S(t) [12,13].

7.5.6.1 Sensitivity Analyses

The following TTF sensitivity analyses will be performed.

- Sensitivity analysis will be conducted
  - 1-ignoring events of treatment discontinuation for reason due to patient preference (subject withdrew consent, subject lost to follow up) or physician withdrawal (subject no longer meets study criteria, administrative reason by sponsor)
  - 2-censoring the TTF at randomization for subjects who were randomized but never treated.
- TTF KM curves by treatment group will be produced for Chinese subjects from the TTF population (HR and 95% CI, median TTF and 95% CI will be presented)

7.5.7 OS Interim Analysis

A formal interim analysis for superiority of OS in subjects who were randomized to nivolumab vs. subjects who were randomized to docetaxel will be performed on all randomized subjects when at

least 291 deaths have been observed (approximately 76% (291/382) of the total number of deaths required for the final analysis).

This OS comparison will be tested using the interim monitoring feature of EAST software based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary to reject H_0, controlling for a two-sided overall α of 5%. For example, if exactly 291 deaths are in the locked database at the interim analysis, H_0 would be rejected if the p-value from the log-rank test is $p < 0.020$. If the number of deaths is not exactly 291 at the time of the interim analysis, the nominal critical point and value of both the interim and final analyses will be calculated based upon the observed information fraction.

The DMC will review the safety and efficacy data from the interim analyses and will determine if the study should continue or should be stopped. If the trial is stopped for superiority at the interim, the p-value will be considered as the final OS result. All secondary endpoint analyses will be tested at that time. The p-values from these analyses will be considered as the final results. See Section 7.5.4 for type I error control for secondary endpoints.

If the study continues beyond the interim analysis and exactly 382 deaths are in the locked database at the final analysis, H_0 would be rejected if the p-value from the log-rank test is $p < 0.044$. All events in the database at the time of the lock will be used. If number of final events exceeds the number specified per protocol, final boundary will not be recalculated using updated information fraction at interim.

7.6 Safety

Safety analyses will be performed by treatment group "as treated". The primary population at theTTF analysis will be the Treated subjects among TTF population. The population at the formal OS analyses will be the all treated population.

7.6.1 Deaths

See Core Safety SAP

7.6.2 Serious Adverse Events

See Core Safety SAP

7.6.3 Adverse Events Leading to Discontinuation of Study Therapy

See Core Safety SAP

7.6.4 Adverse Events Leading to Dose Delay of Study Therapy

See Core Safety SAP

7.6.5 Adverse Events

See Core Safety SAP

7.6.6 Adverse Events by Subgroups

See Core Safety SAP

7.6.7 Multiple Events

See Core Safety SAP

7.6.8 Adverse Events of Special Interest

See Core Safety SAP

7.6.9 Clinical Laboratory Evaluations

The analysis population for each laboratory test is restricted to treated subjects who underwent that laboratory test.

7.6.9.1 Hematology

See Core Safety SAP

7.6.9.2 Serum Chemistry

See Core Safety SAP

7.6.10 Vital Signs and Pulse Oximetry

See Core Safety SAP

7.6.11 Immunogenicity Analysis

Nivolumab arm only. See Core Safety SAP

7.6.12 Pregnancy

By-subject listing of pregnancy tests results will be provided

7.6.13 Clinical Safety Program (CSP)

See Core Safety SAP

7.7 Pharmacokinetics

Pharmacokinetics analyses will be performed on nivolumab arm only.

The concentration vs. time data obtained in this study will be combined with data from other studies in the clinical development program to refine a population PK model. This model will be used to evaluate the effects of ethnic as a categorical covariate and possibly other covaraites on the PK of nivolumab and to determine measures of individual exposure (such as steady-state peak, trough, and time-averaged concentration). Model determined exposures may be used for exposure-response analyses of selected efficacy and safety end points. Results of population PK and exposure-response analyses will be reported separately.

[redacted]

8 CONVENTIONS

The following conventions may be used for imputing partial dates for analyses requiring dates:

- For missing and partial adverse event onset dates, imputation will be performed using the Adverse Event Domain Requirements Specification[19].
- Missing and partial Non-Study Medication Domain dates will be imputed using the derivation algorithm described in 4.3.3 of BMS Non-Study Medication Domain Requirements Specification[20].

For death dates, the following conventions will be used for imputing partial dates:

- If only the day of the month is missing, the 1st of the month will be used to replace the missing day. The imputed date will be compared to the last known date alive and the maximum will be considered as the death date.
- If the month or the year is missing, the death date will be imputed as the last known date alive. If the date is completely missing but the reason for death is present the death date will be imputed as the last known date alive

For date of progression, the following conventions will be used for imputing partial dates:

- If only the day of the month is missing, the 1st of the month will be used to replace the missing day*.
- If the day and month are missing or a date is completely missing, it will be considered as missing.
- In case of the date of death is present and complete, the imputed progression date will be compared to the date of death. The minimum of the imputed progression date and date of death will be considered as the date of progression.
- For other partial/missing dates, the following conventions may be used:
- If only the day of the month is missing, the 15th of the month will be used to replace the missing day.
- If both the day and the month are missing, "July 1" will be used to replace the missing information.
- If a date is completely missing, it will be considered as missing.

The following conversion factors will be used to convert days to months or years: 1 month = 30.4375 days and 1 year = 365.25 days.

Duration (e.g. time from first diagnosis of NSCLC to first dosing date, duration response, and time to response) will be calculated as follows:

Duration = (Last date - first date + 1)

All statistical analyses will be carried out using SAS (Statistical Analysis System software, SAS Institute, North Carolina, USA) unless otherwise noted.

9 CONTENT OF REPORTS

All analyses describe in this SAP will be included in the final Clinical Study Report. All analyses will be repeated for subset of subjects from China for the China CSR. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

10 DOCUMENT HISTORY

Table 10-1: Document History

Version Number	Author(s)	Date	Description
1.0	Christine Baudelet	20Dec2016	Initial version
2.0	Christine Baudelet	01Mar2017	Changes related to Amendment 03 1.2 Schedule of Analyses: timing of Interim and final analysis of OS and TTF IA revised 2.4 Protocol Amendments: amendment 03 added 4.1 Primary Endpoint: statement about crossover rate and treshold for consistency removed 4.2.3 Time to Treatment Failure: definition of TTF endpoint refined. Description of the sensitivity analyses removed as described in section 7.5.6 5 Sample Size and Power: required number of events at OS IA/FA revised, comparison of power with the weighted vs regular log rank test provided, power as function of the cross over rate provided 6.3 Populations for Analyses: TTF population: description revised 7.5.1.1 Primary Analysis of OS: 2 step hierarchical testing added at OS IA. Weighted log rank test added. 7.5.1.2 OS Sensitivity Analyses: regular log-rank test added, sensitivity analysis limited to required number of events for OS IA: number of events revised 7.5.6 TTF Interim Analysis: number of randomized patient to be included in TTF interim analysis revised. DMC review added. 7.5.6.2 TTF Sensitivity Analyses: section revised 7.5.7 OS Interim Analysis: required number of events at OS IA/FA revised Other changes: Table 4.2.2-1: Censoring Scheme for Primary Definition of PFS: censoring rule when no baseline tumor assessment revised (no death condition removed) 7.5.1.8 OS Analysis at time of TTF interim analysis: analysis by histology added
2.1	Christine Baudelet	02May2017	Changes related to Protocol Amendment 04 Analyses of OS/PFS for TTF population at time of TTF interim analysis removed Definition of TTF endpoint refined