Statistical Analysis Plan for

Official Title of Study

A PHASE 3 RANDOMIZED, DOUBLE-BLIND, MULTI-CENTER STUDY OF ADJUVANT NIVOLUMAB VERSUS PLACEBO IN SUBJECTS WITH HIGH RISK INVASIVE UROTHELIAL CARCINOMA

NCT02632409

16-November-2019

STATISTICAL ANALYSIS PLAN

A PHASE 3 RANDOMIZED, DOUBLE-BLIND, MULTI-CENTER STUDY OF ADJUVANT NIVOLUMAB VERSUS PLACEBO IN SUBJECTS WITH HIGH RISK INVASIVE UROTHELIAL CARCINOMA

PROTOCOL CA209-274

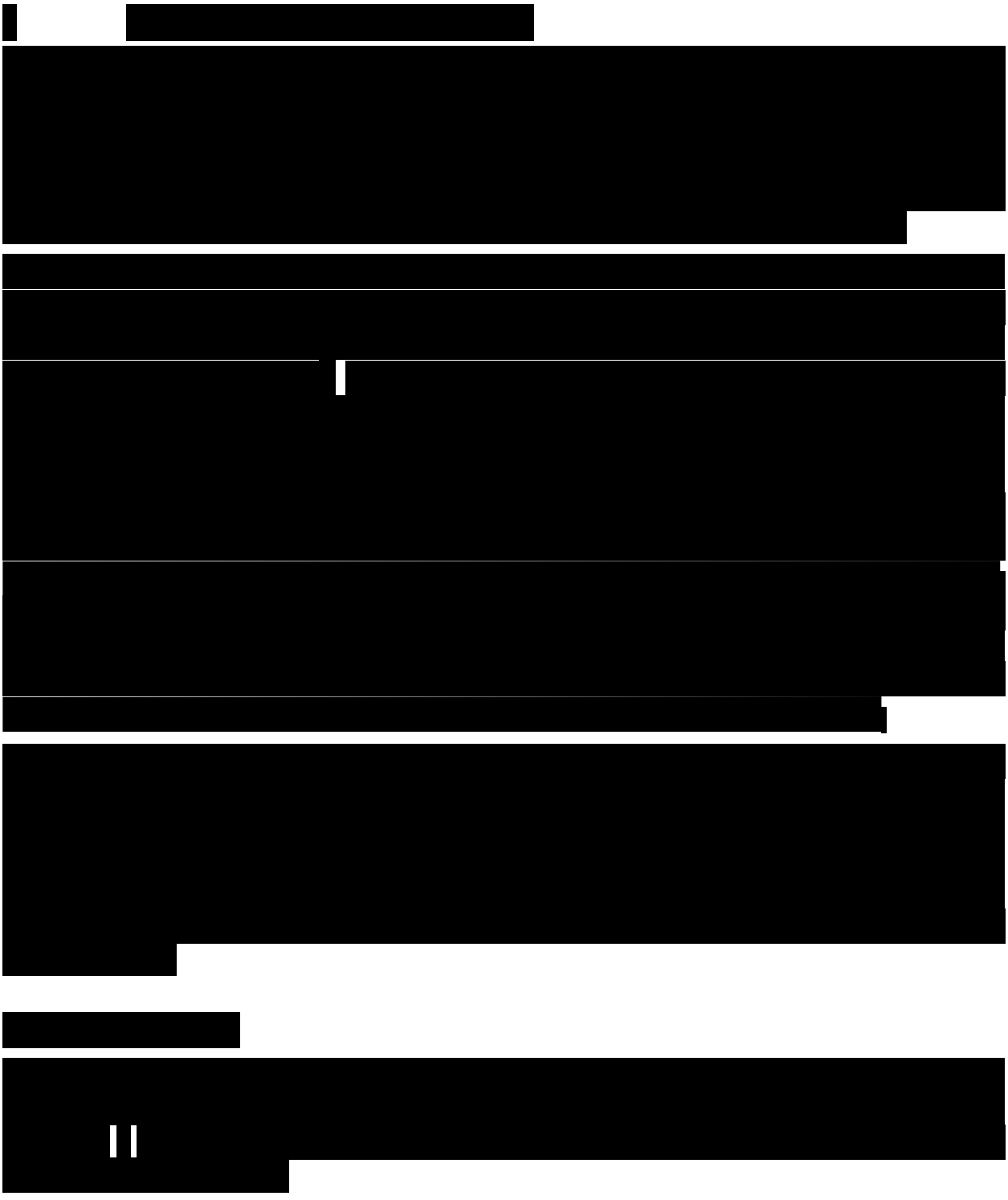
VERSION # 2.0

DATE: 06-NOV-2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

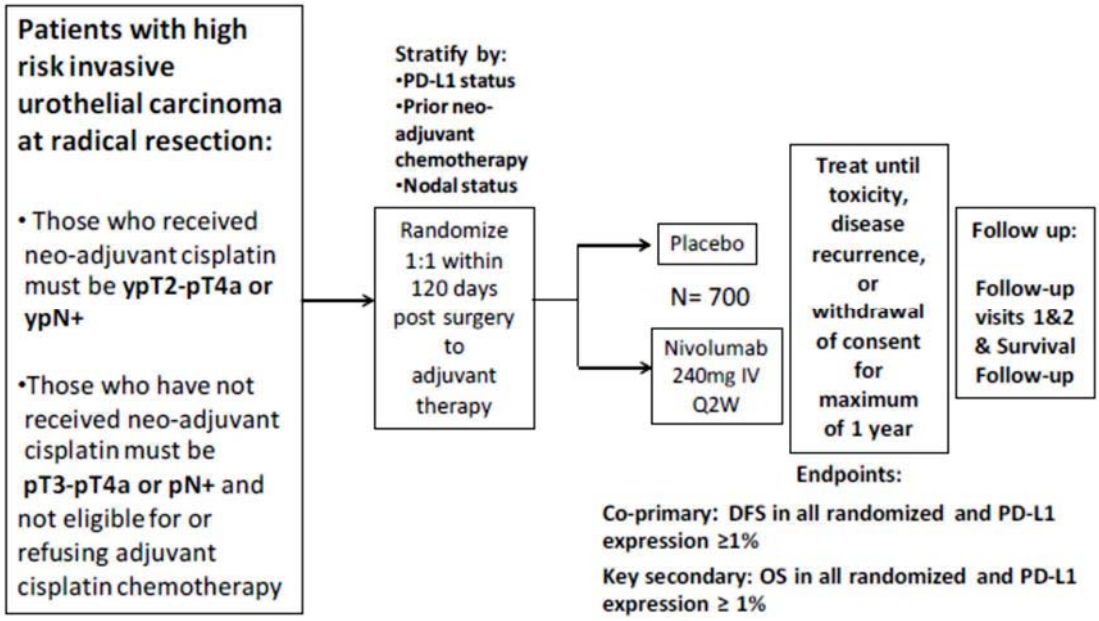


2 STUDY DESCRIPTION

2.1 Study Design

The study design schematic is presented in Figure 2.1-1.

Figure 2.1-1: Study Design Schematic



This is a phase 3, randomized, double-blind, multicenter trial of adjuvant nivolumab versus placebo in adult male or female subjects who have undergone radical resection of IUC originating in the bladder or upper urinary tract (renal pelvis or ureter) and are at high risk of recurrence:

- Subjects who received neo-adjuvant cisplatin chemotherapy: ypT2-pT4a or ypN+
- Subjects who have not received neo-adjuvant cisplatin chemotherapy: pT3-pT4a or pN+ and are not eligible for or refusing adjuvant cisplatin chemotherapy

Approximately 700 subjects will be randomized in a blinded fashion 1:1 to nivolumab versus placebo within 120 days of radical resection and stratified by nodal status (N+ vs. N0/x with < 10 nodes removed vs. N0 with ≥ 10 nodes removed), tumor PD-L1 expression (≥ 1%, < 1% / indeterminate), and use of cisplatin neo-adjuvant chemotherapy.

Treatment, in the absence of prohibitive toxicities, disease recurrence/progression, or withdrawal of consent will be continued for a maximum of 1 year. The co-primary endpoint is DFS in subjects with tumors expressing PD-L1 at ≥ 1% (2.5% alpha) and in all randomized subjects (2.5% alpha). The overall sample size is set up to allow a clinically meaningful effect to be statistically significant at alpha level of 2.5% (two-sided) in the all randomized group.

Even weighting of the alpha distribution between DFS assessment in all randomized (N = 700) and PD-L1 expressers (~ 42% of the all randomized population) reflects an assumption of enrichment of nivolumab efficacy in the PD-L1 expressers.

At the time of the revised protocol 03, the anticipated prevalence of PD-L1 expression ≥ 1% was 46% (original study design anticipated 50% prevalence) and the PD-L1 expression < 1% population in the study was capped at 54%.

During the execution of the study the PD-L1 at ≥ 1% rate was ≈ 42%, for this reason in revised protocol 04 the cap was removed to make the study sample representative of the study population.

Hence, it is anticipated that the final population will include approximately 42% of patients who are PD-L1 expression ≥ 1%.

The number of randomized subjects with upper tract urothelial cancers (UTUC) is capped at approximately 20% (140 subjects) of total global enrollment. Once approximately 140 subjects with UTUC are randomized, only subjects with bladder cancer are enrolled.

Following discontinuation of study therapy, subjects will be followed for survival and those that have not had a non-urothelial tract recurrence will be followed for recurrence.

2.2 Treatment Assignment

After informed consent has been obtained and the subject's eligibility is established, the subject will be enrolled and a number will be assigned through an interactive voice response system (IVRS). Subjects will be randomized in a 1:1 ratio to Arm A (nivolumab) or Arm B (placebo) stratified by the following factors:

- PD-L1 expression level (≥ 1% vs. < 1%/indeterminate)
- Receipt of neo-adjuvant cisplatin based chemotherapy for IUC (Yes/No)
- Pathologic status of disease in lymph nodes (N+ vs. N0/x with < 10 nodes removed vs. N0 with ≥ 10 nodes removed)

2.3 Blinding and Unblinding

The Sponsor, subjects, investigator and site staff will be blinded to the study therapy administered. Each investigative site must assign an unblinded pharmacist/designee, and an unblinded site monitor will be assigned to provide oversight of drug supply and other unblinded study documentation.

2.4 Protocol Amendments

This SAP incorporates the following amendments:

Table 2.4-1: Protocol Amendments

Amendments	Date of Issue	Summary of Major Changes
Revised Protocol 1	21-OCT-2015	Modified randomization rate from 2:1 to 1:1;increased total number of treated subjects from 600 to 640; other minor changes, including simplifying the description of the study population and definition of recurrences
Revised Protocol 2	18-AUG-2016	Extends acceptable period of time between radical resection and randomization from 90 days to 120 days; clarifies pathology language for eligibility; clarifies recurrence language, updates safety and contraceptive language to be consistent with nivolumab Investigators Brochure version 15; other minor changes.
Revised Protocol 3	18-JUL-2017	Provides a more precise description of low- and high-risk non-muscle invasive bladder cancer (NMIBC) conforming to nomenclature used in the 1973 WHO and 2004 WHO/ISUP guidelines; aligns the synopsis with the body of the protocol with respect to non-eligibility of subjects with high-

Table 2.4-1: Protocol Amendments

Amendments	Date of Issue	Summary of Major Changes
		risk NMIBC; states that subjects with CIS at urethral or ureteral surgical margins are not eligible for study; changes the PD-L1 expression ≥ 1% from 50% to 46% in accord with recently published data from the CA209-275 registrational study; applies a 20% cap on the number of pelvis and ureter cancer subjects randomized to the study; states that suspect lesions observed during screening procedures must be discussed with the MM prior to randomization; other minor changes.
Revised Protocol 4	19-DEC-2018	Incorporates statistical modifications (including sample size increases from 640 to 700), removal of the cap on PD-L1 < 1%, and adds PFS2 as exploratory endpoint. Revisions to Inclusion and Exclusion Criteria; and timing of TSH laboratory tests. Updates to the most recent guidance for nivolumab treatment. Minor formatting and typographical corrections were corrected. [redacted]
Revised Protocol 5	XX-SEP-2019	Addition of a formal interim OS analysis at the time of final DFS analysis for both study populations (PDL1 ≥ 1% and all randomized subjects). [redacted]

2.5 Data Monitoring and Other External Committees

A Data Monitoring Committee (DMC) will be established to provide oversight of safety and efficacy considerations in protocol CA209274. Additionally, the DMC will provide advice to the sponsor regarding actions the committee deems necessary for the continuing protection of subjects enrolled in the study. The DMC will be charged with assessing such actions in light of an acceptable benefit/risk profile for nivolumab. The DMC will act in an advisory capacity to BMS and will monitor subject safety and evaluate the available efficacy data for the study. The oncology therapeutic area of BMS has primary responsibility for design and conduct of the study.

3 OBJECTIVES

3.1 Co-Primary Objectives

- To compare the DFS for nivolumab versus placebo in all randomized subjects
- To compare the DFS for nivolumab versus placebo in subjects with tumors expressing PD-L1 (≥ 1% membranous staining in tumor cells)

3.2 Secondary Objectives

- To compare overall survival (OS) for nivolumab versus placebo in subjects with tumors expressing PD-L1 (≥ 1% membranous staining in tumor cells) and all randomized subjects
- To evaluate non-urothelial tract recurrence free survival (NUTRFS) for nivolumab versus placebo in subjects with tumors expressing PD-L1 (≥ 1% membranous staining in tumor cells) and all randomized subjects

- To evaluate the disease survival specific (DSS) for nivolumab and placebo in subjects with tumors expressing PD-L1 (≥ 1% membranous staining in tumor cells) and all randomized subjects

[redacted]

4 ENDPOINTS

4.1 Efficacy Endpoints

4.1.1 Disease-Free Survival

Disease-Free Survival is the primary endpoint of the study and will be programmatically determined based on the disease recurrence date provided by the investigator.

The following recurrence types are included in the DFS endpoint:

- Local, urothelial tract: any high- and intermediate-risk NMIBC and any new invasive urothelial carcinoma in the lower or upper urothelial tract (defined as T2 or greater), including lesions thought to be a second urothelial carcinoma primary
- Local, non-urothelial tract: Any recurrence in pelvic soft tissue or involving pelvic nodes below the aortic bifurcation
- Distant: Any non-local recurrence

Low-risk NMIBC will **not** be reported as a recurrence event for the DFS endpoint.

Subjects who die without a reported recurrence will be considered to have recurred on the date of their death.

DFS rate at time *T* (e.g. 6 months, depending on the minimum follow-up) is defined as the probability that a subject has no disease recurrence and is alive at time *T* following randomization, based on the Kaplan-Meier estimate.

The first on-study non-cystoscopy tumor assessment is scheduled to be conducted at 12 weeks (± 1 week) from the date of first dose. Subsequent non-cystoscopy tumor assessments are scheduled every 12 weeks (± 1 week) up to Week 96, then every 16 weeks (± 2 weeks) up to Week 160, then every 24 weeks (± 4 weeks) until non-urothelial tract recurrence or treatment is discontinued (whichever occurs later) for a maximum of 5 years.

Cystoscopy in subjects with upper GU primaries who still have bladder intact will occur in addition to other tumor imaging assessments every 12 weeks (± 1 week) from the date of first dose to Week 48, then every 24 weeks (± 2 weeks) up to Week 96, then every 48 weeks (± 4 weeks) until non-urothelial tract recurrence or treatment is discontinued (whichever occurs later) for a maximum of 5 years.

The **primary definition of DFS (DFS truncated at subsequent therapy and new non-urothelial carcinoma primary cancer)** is defined as the time between the date of randomization and the date of first documented recurrence (local urothelial tract, local non urothelial tract or distant), or death due to any cause, whichever occurs first. The primary definition of DFS accounts for subsequent anticancer therapy (see Table 4.1.1-1[1)b)]) and new non-urothelial carcinoma primary cancer by censoring at the last evaluable disease assessment on or prior to the date of subsequent therapy/new non-urothelial carcinoma primary cancer.

Detailed censoring rules for the primary definition of DFS are presented in Figure 4.1.1-1 and Table 4.1.1-1.

The secondary definition of DFS accounts for disease assessments occurring on or after initiation of subsequent anticancer therapy. Censoring scheme will be the same as for the primary DFS definition except that new anticancer therapy censoring that will be ignored in this sensitivity analysis. The secondary definition of DFS accounts for new non-urothelial carcinoma primary cancer by censoring at the last evaluable disease assessment on or prior to the date of new non-urothelial carcinoma primary cancer.

Figure 4.1.1-1: DFS Primary Definition

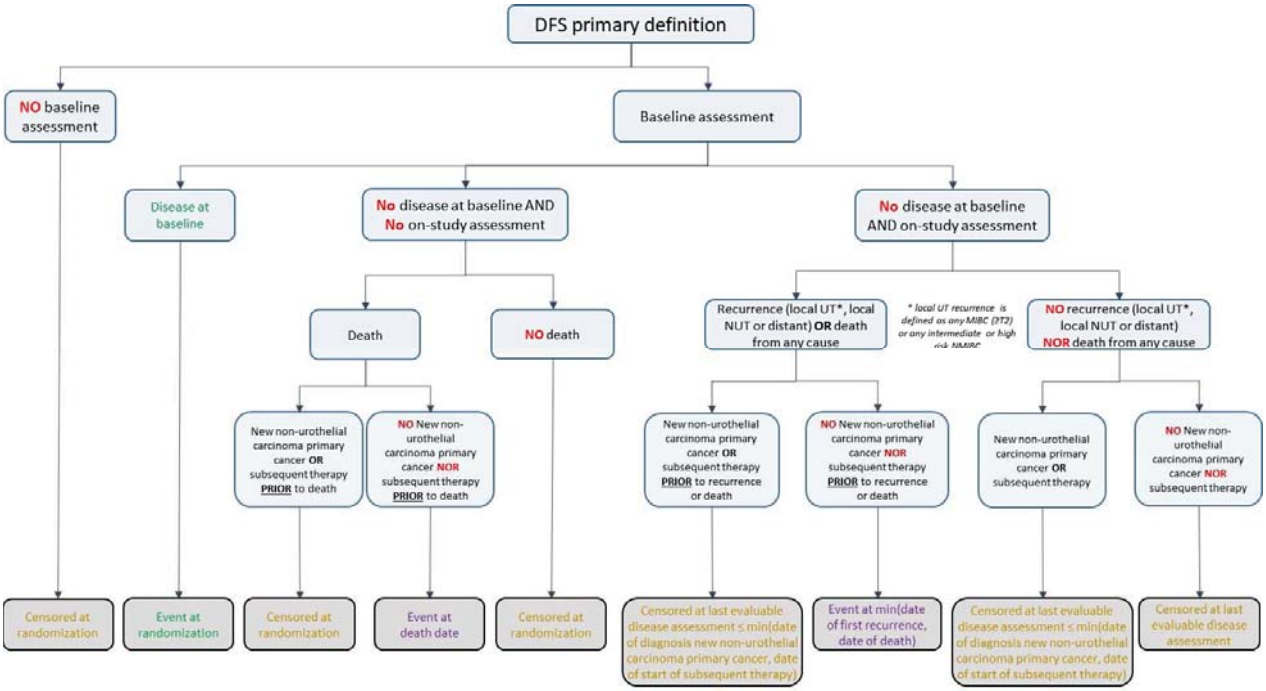


Table 4.1.1-1: Censoring Scheme used in Primary Definition of DFS

Situation	Date of Progression or Censoring	Outcome
No baseline disease assessments	Date of randomization	Censored
No on study disease assessments and no death[1)a)]	Date of randomization	Censored
No recurrence and no death, and no new anticancer therapy started[1)b)] nor new non-urothelial carcinoma primary cancer	Date of last evaluable disease assessment	Censored
Disease at baseline	Date of randomization	DFS event
Any urothelial carcinoma recurrence (i.e. local urothelial tract, local non urothelial tract or distant)	Date of the first recurrence	DFS event

Table 4.1.1-1: Censoring Scheme used in Primary Definition of DFS

Situation	Date of Progression or Censoring	Outcome
Death from any cause without recurrence and no new anticancer therapy started[1)b)] nor new non-urothelial carcinoma primary cancer	Date of death	DFS event
New anticancer therapy1)b) received without recurrence reported prior to or on the same day	Date of last evaluable disease assessment prior to or on the same date of initiation of subsequent therapy	Censored
New non-urothelial carcinoma primary cancer without urothelial carcinoma recurrence reported prior to or on the same day	Date of last evaluable disease assessment prior to or on the same date of diagnosis of new non-urothelial carcinoma primary cancer	Censored

a) Disease assessments and death if any, occurring after start of subsequent anti-cancer therapy or after new non-urothelial carcinoma primary cancer are not considered.

b) New anticancer therapy includes systemic therapy, tumor directed radiotherapy and tumor directed surgery (**except TURBT**). As per protocol, the only exception is that a single dose of intravesical chemotherapy instilled immediately after resection of low-risk NMIBC (not considered as a DFS event), is permitted. Therefore **any use of one single dose of chemotherapy given via intravesical route should NOT be considered as subsequent treatment**.

4.1.2 Overall Survival

Overall survival (OS) is key secondary endpoint and defined as the time from randomization to the date of death from any cause. For subjects that are alive, their survival time will be censored at the date of last contact date (or "last known alive date"). OS will be censored at the date of randomization for subjects who were randomized but had no follow-up.

Survival follow-up will be conducted every 3 months after subject's off-treatment date.

4.1.3 Non-Urothelial Tract Recurrence-Free Survival

Non-Urothelial Tract Recurrence-Free Survival (NUTRFS) is a secondary endpoint of the study and will be programmatically determined based on the disease recurrence date provided by the investigator.

The following NUT recurrences are included in the NUTRFS endpoint:

- Local, Non-Urothelial Tract: Any recurrence in pelvic soft tissue or involving pelvic nodes below the aortic bifurcation
- Distant: Any non-local recurrence

Local urothelial tract recurrence (defined as low-, intermediate- or high-risk NMIBC or invasive urothelial carcinoma in the lower or upper urothelial tract including lesions thought to be a second urothelial carcinoma primary) will **not** be reported as a recurrence event for the NUTRFS endpoint.

Subjects who die without a reported recurrence will be considered to have recurred on the date of their death.

NUTRFS rate at time *T* (e.g. 6 months, depending on the minimum follow-up) is defined as the probability that a subject has no local non-urothelial tract nor distant recurrence and is alive at time *T* following randomization, based on Kaplan-Meier estimate.

As for DFS, the primary definition of NUTRFS (NUTRFS truncated at subsequent therapy **and new non-urothelial carcinoma primary cancer**) is defined as the time between the date of randomization and the date of first documented recurrence (local non-urothelial tract or distant), or death due to any cause, whichever occurs first. The primary definition of NUTRFS accounts for subsequent therapy and new non-urothelial carcinoma primary cancer by censoring at the last evaluable disease assessment on or prior to the date of subsequent therapy/new non-urothelial carcinoma primary cancer.

Detailed censoring rules for the primary definition of NUTRFS are presented in Table 4.1.3-1 and Figure 4.1.3-1.

Figure 4.1.3-1: NUTRFS Primary Definition

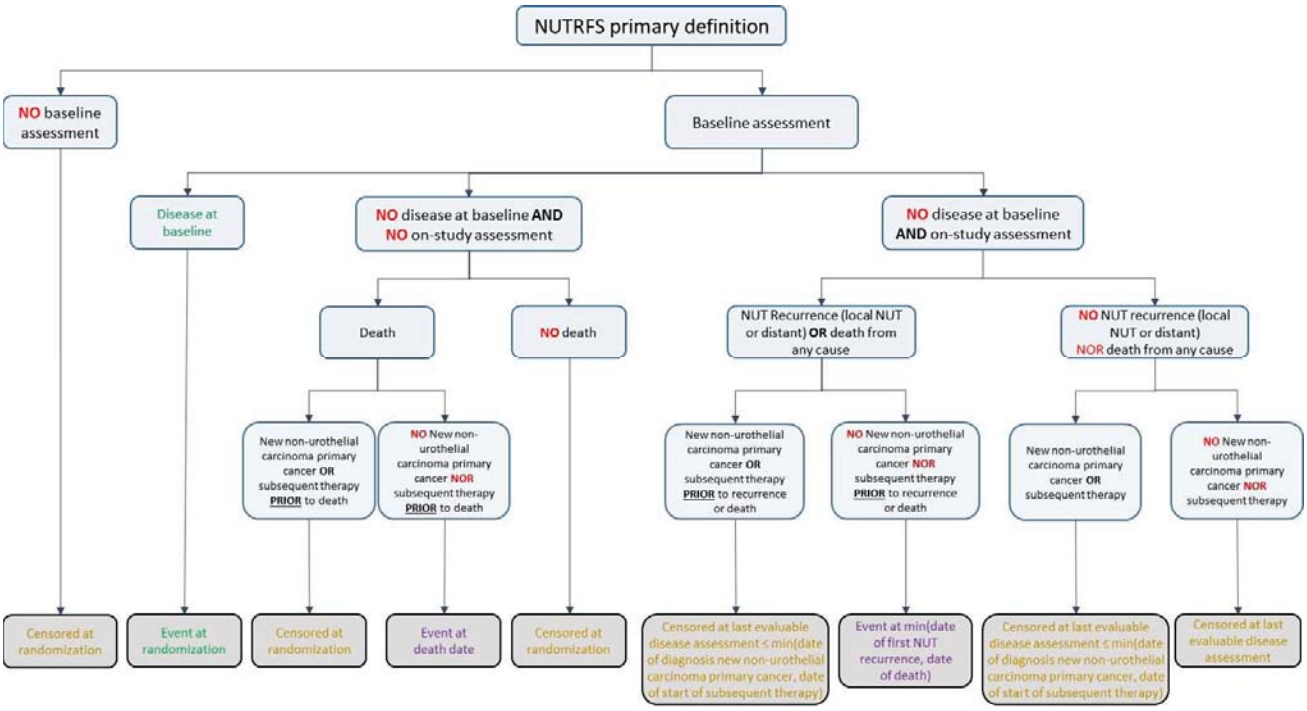


Table 4.1.3-1: Censoring Scheme used in Primary Definition of NUTRFS

Situation	Date of Progression or Censoring	Outcome
No baseline disease assessments	Date of randomization	Censored
No on study disease assessments and no death[c)]	Date of randomization	Censored
No NUT recurrence and no death, and no new anticancer started[d)] nor new non-urothelial carcinoma primary cancer	Date of last evaluable disease assessment	Censored
Disease at baseline	Date of randomization	NUTRFS event
NUT recurrence (i.e. local non-urothelial tract recurrence or distant)	Date of the first NUT recurrence	NUTRFS event
Death from any cause without NUT recurrence and no new anticancer started[d)] nor new non-urothelial carcinoma primary cancer	Date of death	NUTRFS event
New anticancer therapy[d)] received without NUT recurrence reported prior to or on the same day	Date of last evaluable disease assessment prior to or on the same date of initiation of subsequent therapy	Censored
New non-urothelial carcinoma primary cancer without NUT recurrence reported prior to or on the same day	Date of last evaluable disease assessment prior to or on the same date of diagnosis of new non-urothelial carcinoma primary cancer	Censored

c) Disease assessments and death if any, occurring after start of subsequent anti-cancer therapy or after new non-urothelial carcinoma primary cancer are not considered.

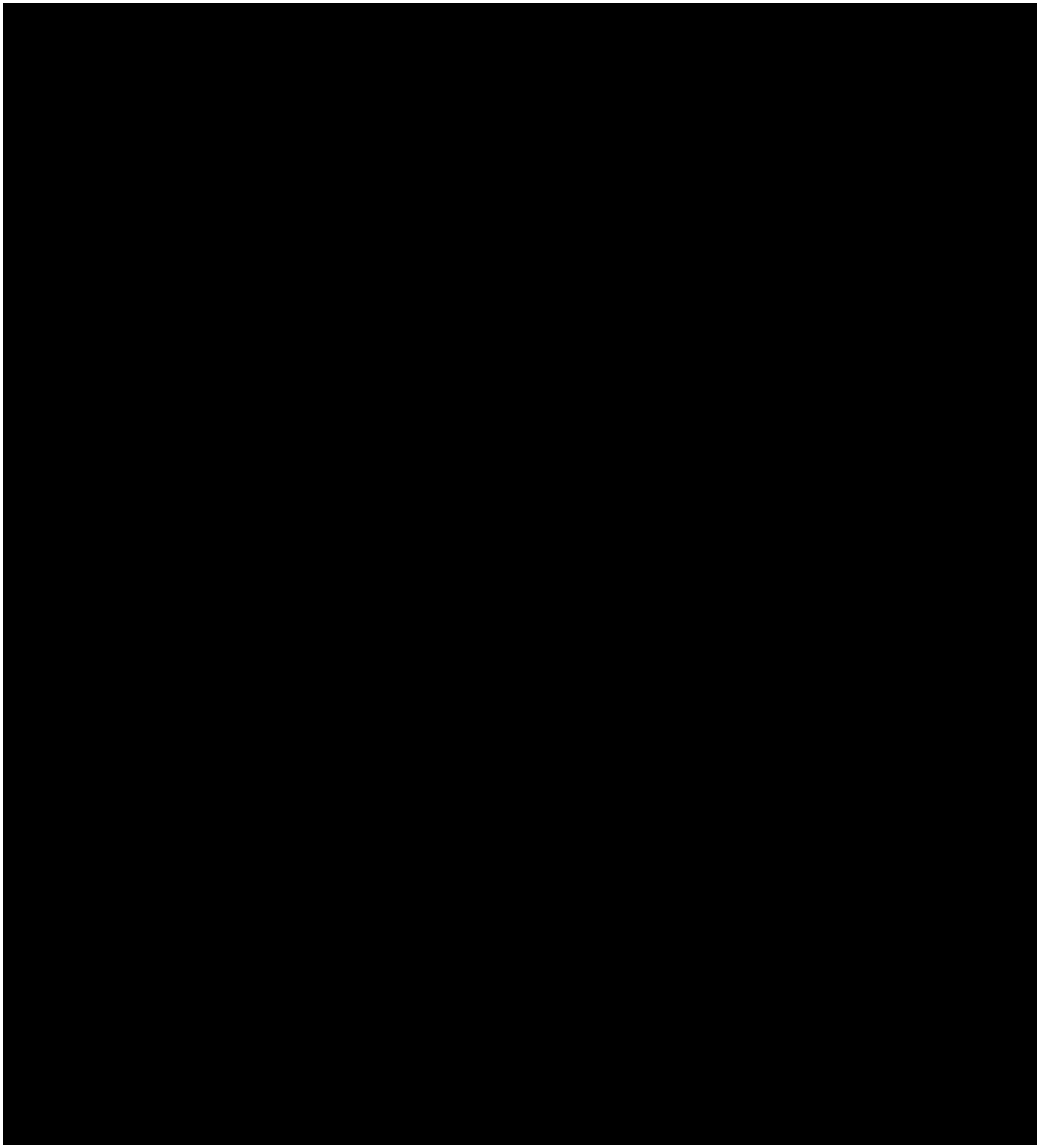
d) New anticancer therapy includes systemic therapy, tumor directed radiotherapy and tumor directed surgery (except TURBT). As per protocol, the only exception is that a single dose of intravesical chemotherapy instilled immediately after resection of low-risk NMIBC (not considered as a DFS event), is permitted. Therefore any use of chemotherapy given via intravesical route (including "BCG") should NOT be considered as subsequent treatment.

4.1.4 Disease-Specific Survival

Disease-Specific survival (DSS) is a secondary endpoint and defined as the time from randomization to the date of death due to disease (urothelial cancer). For subjects that are alive, their survival time will be censored at the date of last contact date (or "last known alive date"). DSS will be censored at the date of randomization for subjects who were randomized but had no follow-up.

Death not due to disease will be considered as a competing risk. Consequently, for subjects who died due to other cause than urothelial cancer, DSS in the analysis of cause-specific hazard analysis will be censored on the date of death.

DSS will be followed continuously as part of OS follow-up.



5 SAMPLE SIZE AND POWER

The sample size for this study is based on a comparison of the DFS distribution between subjects randomized to receive nivolumab and subjects randomized to receive placebo. There will be two co-primary comparisons: in subjects with PD-L1 expression ≥ 1% and in all randomized subjects with an alpha allocation of 2.5% (two-sided) each.

The prevalence of subjects with PD-L1 expression ≥ 1% is expected to be around 42%. In time-to-events trials, the number of events and power may typically be calculated assuming an exponential distribution in each treatment arm. However for DFS, a meaningful number of long-term disease-free survivors may be observed. As a consequence, the DFS curves may not follow an exponential decay and a flattening of the curves may be observed toward the end of the tail. Consequently, an exponential cure rate distribution was used to calculate the sample size.

The exponential cure rate distribution can be described with following formula:

$$S(t) = p + (1 - p) * \{e^{-\alpha t}\},$$

where p is the cure rate and α is the parameter of the exponential distribution for the non-cured population.

Based on historical data[1], median DFS distribution in the control arm is assumed to be 12 months with a 5 year DFS rate of 32% used as the cure rate (i.e., p=0.32 and α=0.11). It is also assumed that DFS distribution in control arm does not depend on the PD-L1 expression level. Cure rates in subjects with PD-L1 expression ≥ 1% and in the all randomized population for the experimental arm are targeted to be 54% and 47% respectively while non cured population is assumed to follow same distribution as in the control arm. It is assumed that the effect of nivolumab vs. placebo will be delayed. A study of nivolumab vs. ipilimumab in metastatic melanoma (CheckMate 238[12]) suggested a 3-month delayed effect for nivolumab, therefore the model used for the experimental arm also assumes a 3-month delayed effect in subjects with PD-L1 expression ≥ 1% and in the all randomized population for the experimental arm.

Approximately 700 subjects will be randomized in 1:1 ratio to nivolumab arm and control arm respectively.

- Approximately 410 DFS events in all randomized subjects will provide around 87% power to detect an average HR of 0.72 with an overall type I error of 2.5% (two-sided)
- Approximately 162 DFS events in subjects with PD-L1 expression ≥ 1% (~294 subjects) will provide around 80% power to detect an average HR of 0.61 with an overall type I error of 2.5% (two-sided).

The target average hazard ratios were obtained via simulation using above modeling.

It is expected that the required number of events in each population will be reached when all subjects will have at least a minimum of 17 months of follow-up.

An interim DFS analysis (in all randomized subjects and in subjects with PD-L1 expression ≥ 1%) will also be performed when 85% of DFS events in all randomized subjects (i.e., 348 DFS events) and in subjects with PD-L1 expression ≥ 1% (i.e., 137 DFS events) will be observed. The alpha level for DFS will be adjusted for the planned interim analysis using Lan-DeMets alpha spending function with the O'Brien-Fleming type of boundary in East v6.

Assuming an average accrual rate of 16 subjects per month, the accrual will take approximately 43 months. The total duration of the study from start of randomization to final analysis of DFS is expected to be 60 months (43 months of accrual + 17 months of follow-up).

DFS analyses in all randomized population and in subjects with PD-L1 expression ≥ 1% are projected to occur 45 months after start of the study for the interim analyses and 60 months (43 months of accrual + 17 months of follow-up) for the final analyses (Table 5-1 and Table 5-2).

It is expected that interim DFS analysis (defined as analysis timepoint T_1) in each population will occur at the same time. However, because the DFS event rate pattern may be different between all randomized population and subjects with PD-L1 expression ≥ 1%, the interim DFS analysis of each population might also occur separately.

The final analysis of DFS in each population will be performed at the same time when 410 and 162 events of DFS are observed in all randomized subjects and in subjects with PD-L1 expression ≥ 1%, respectively (defined as analysis timepoint T_2).

Table 5-1: Schedule of Analyses of Primary Endpoint in all Randomized Population

	Interim DFS Analysis	**Final DFS Analysis**
Population	**All Randomized Subjects**	
Conditions	**When 348 Events of DFS are observed in all Randomized Subjects**	**When 410 Events of DFS are observed in all Randomized Subjects**
Expected Timing	45 months	60 months (43 months of accrual + 17 months of follow-up)
Nominal Significance Level/Power	0.01349149	0.02108499 / 87%
Lower Boundary for Statistical Significance	Observed hazard ratio of 0.76	Observed hazard ratio of 0.79

Table 5-2: Schedule of Analyses of Primary Endpoint in Subjects with PD-L1 Expression ≥ 1%

	Interim DFS Analysis	Final DFS Analysis
Population	**Subjects with PD-L1 Expression ≥ 1%**	
Conditions	When 137 events of DFS are observed in subjects with PD-L1 expression ≥ 1%	When 162 events of DFS are observed in subjects with PD-L1 expression ≥ 1%
Expected Timing	45 months	60 months (43 months of accrual + 17 months of follow-up)
Nominal Significance Level/Power	0.01349149	0.02108499 / 80%
Lower Boundary for Statistical Significance	Observed hazard ratio of 0.65	Observed hazard ratio of 0.69

Key secondary endpoint of OS will be formally compared (in subjects with PD-L1 expression ≥ 1% and all randomized subjects) using a hierarchical procedure in each population.

5.1 Key OS secondary Endpoint Analyses

The secondary endpoint of OS will be formally compared (in subjects with PD-L1 expression ≥ 1% and all randomized subjects) using a hierarchical procedure in each population as discussed in Glimm et al.[16] with overall alpha of 2.5% (two-sided) using Lan-DeMets alpha spending function with the O'Brien-Fleming type of boundary in East v6 for each endpoint.

In all randomized subjects

OS could be formally compared either

1) at the final analysis for DFS (T2) that will occur after 410 events of DFS are observed or at approximately 230 OS events observed, if no final DFS analysis is needed.
2) after 303 OS events have been observed (T3).
3) after 404 OS events have been observed (T4).

This testing procedure is also summarized in Figure 5.1-1.

Approximately 404 OS events in all randomized subjects will provide approximately 87% power to detect an HR of 0.71 with an overall type I error of 2.5% (two-sided). This corresponds to a median increase from 4.6 years to 6.5 years under exponential assumption. Two interim analyses are planned for OS in all randomized subjects at ~57% (i.e., 230 events) and at 75% (i.e., 303 events) information fractions.

Figure 5.1-1: Hierarchical Procedure with Group Sequential Testing in all Randomized Subjects

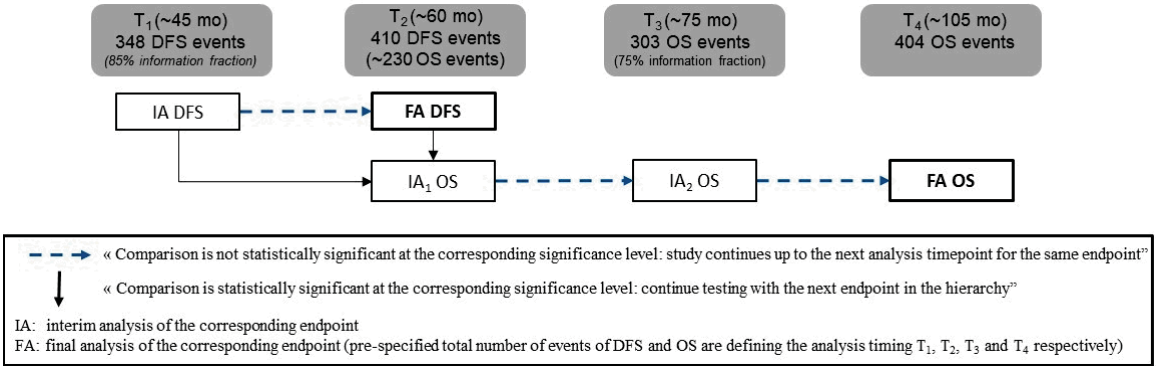


Table 5-1.1: Schedule of Analyses of OS, Key Secondary Endpoint, in all Randomized-Population

	Interim OS Analysis #1	Interim OS Analysis #2	Final OS Analysis
Population	**All Randomized Subjects**		
Conditions	At final DFS analysis (ie approximately 230 OS events should be observed) OR when 230 OS events are observed, in case of no final DFS analysis	When 303 OS events are observed	When 404 OS events are observed
Expected Timing	60 months (T_2)	75 months (T_3)	105 months (T_4)
Nominal Significance Level/Power	0.001877	0.007260	0.022476 / 87%
Lower Boundary for Statistical Significance	Observed hazard ratio of 0.66	Observed hazard ratio of 0.73	Observed hazard ratio of 0.79

In subjects with PD-L1 expression ≥ 1%

Similar procedure (with 2 interim OS analyses) will applied for the secondary endpoint comparisons in subjects with PD-L1 expression ≥ 1% and is summarized in Figure 5.1-2.

OS could be formally compared:

1) at the final analysis of DFS (defined as T2) that will occur after 162 DFS events have been observed or at approximately 91 OS events observed, if no final DFS analysis is needed.
2) after 125 OS events have been observed (defined as T3).
3) after 166 OS events have been observed (defined as T4).

Approximately 166 OS events in subjects with PD-L1 expression ≥ 1% will provide approximately 82% power to detect an HR of 0.61 with an overall type I error of 2.5% (two-sided). This corresponds to a median increase from 4.6 years to 7.5 years under exponential assumption. Two interim analyses are planned for OS in all randomized subjects at ~55% (i.e., 91 events) and at 75% (i.e., 125 events) information fractions.

Figure 5.1-2: Hierarchical Procedure with Group Sequential Testing in the Subjects with PD-L1 Expression ≥ 1%

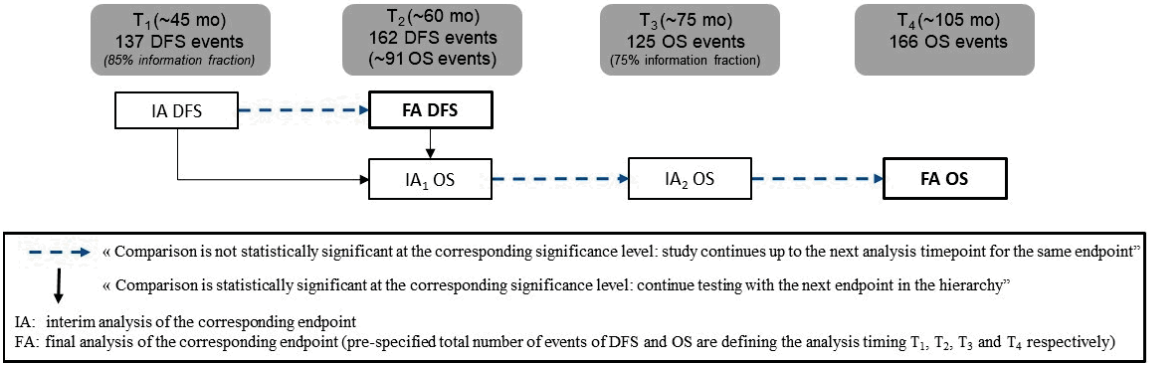


Table 5-1.2: Schedule of Analyses of OS, Key Secondary Endpoint, in the subjects with PD-L1 expression ≥ 1%

	Interim OS analysis #1	Interim OS analysis #2	Final OS analysis
Population	**Subjects with PD-L1 expression level ≥ 1%**		
Conditions	**At final DFS analysis (i.e. approximately 91 OS events should be observed) OR when 91 OS events are observed in subjects with PD-L1 expression level ≥ 1%, in case of no final DFS analysis**	**When 125 OS events are observed in subjects with PD-L1 expression level ≥ 1%**	**When 166 OS events are observed in subjects with PD-L1 expression level ≥ 1%**
Expected timing	60 months (T_2)	75 months (T_3)	105 months (T_4)
Nominal significance level/Power	0.001515	0.007368	0.022494 / 82%
Lower boundary for statistical significance	Observed hazard ratio of 0.51	Observed hazard ratio of 0.61	Observed hazard ratio of 0.70

The interim and final DFS analysis are projected to occur around 45 and 60 months after the start of accrual (T_1, T_2), respectively followed by T_3 (interim analysis of OS) around 75 months and T_4 (final analysis of OS) around 105 months after the start of accrual.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

- Baseline period:
  - Baseline evaluations or events will be defined as evaluations or events that occur before the date and time of the first dose of study treatment. Evaluations (laboratory tests and vital signs) on the same date and time of the first dose of study treatment will be considered as baseline evaluations. Adverse Events (AEs) on the same date and time of the first dose of study treatment will not be considered as pre-treatment events.
  - In cases where the time (onset time of event or evaluation time and dosing time) is missing or not collected, the following definitions will apply:
    - Pre-treatment AEs will be defined as AEs with an onset date prior to but not including the day of the first dose of study treatment;
    - Baseline evaluations (laboratory tests and vital signs) will be defined as evaluations with a date on or prior to the day of first dose of study treatment.
  - If there are multiple valid assessments on or prior to the first dose of study treatment:
    - [redacted]
    - [redacted]
    - For PD-L1, among the biopsy samples records taken prior to or on first dose date (and time if collected), identify first those with quantifiable test result. If there are no records with quantifiable test result, then select those with indeterminant result ("INDETERMINATE"). If there are no records with indeterminant test result, then select those with unavailable result ("NOT EVALUABLE). If there are no records with unavailable test result, then select those with not reported or not available result (all other records). The latest record will be used as the baseline in the analyses. If there is more than one record for the latest date, then choose the one with the greatest specimen ID.
    - [redacted]

- Post baseline period:
  - On-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of study treatment (or with an onset date on or after the day of first dose of study treatment if time is not collected or is missing). For subjects who are off study treatment, AEs will be included if event occurred within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment. No "subtracting rule" will be applied when an AE occurs both pre-treatment and post-treatment with the same preferred term and grade.
  - On-treatment evaluations (laboratory tests and vital signs) will be defined as evaluations taken after the day (and time, if collected and not missing) of first dose of study treatment. For subjects who are off study treatment, evaluations should be within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of study treatment.
- Late-emergent drug-related AEs will be defined as drug-related AEs with an onset date greater than 100 days after the last dose of study treatment in subjects who are off study treatment

6.2 Treatment Regimens

Treatment group "**as randomized**" corresponds to the treatment group assigned by the Interactive Response Technology (IRT) system.

The treatment group "**as treated**" will be same as the treatment group "as randomized" by IRT unless a subject received the **incorrect study treatment for the entire period of treatment**, in which case the subject's treatment group "as treated" will be defined as the incorrect study treatment.

Unless otherwise specified, the safety analysis will be based on the treatment group "**as treated**".

Unless otherwise specified, the efficacy analysis will be based on the treatment group "**as randomized**".

6.3 Populations for Analyses

- Enrolled subjects: All subjects who signed the informed consent form and obtained a subject number.
- Randomized subjects: All subjects who were randomized through the IRT.
- Randomized subjects with PD-L1 expression ≥ 1%: All subjects who were randomized through the IRT with baseline PD-L1 expression ≥ 1% as stratification factor (source IVRS).
- Treated subjects: All randomized subjects who received at least one dose of any study treatment (nivolumab or placebo).
- Treated subjects with PD-L1 expression ≥ 1%: All randomized subjects with baseline PD-L1 expression ≥ 1% as stratification factor (source IVRS) who received at least one dose of any study treatment (nivolumab or placebo).

[redacted]

[redacted]

[redacted]

Analyses of the patients enrolled into the study but not randomized and the reason for not being randomized will be performed on the dataset of enrolled subjects.

Unless otherwise specified, the safety analyses will be analyzed in the treated subjects and in the treated subjects with PD-L1 expression ≥ 1% populations.

Unless otherwise specified, the dosing analyses will be analyzed in the treated subjects and in the treated subjects with PD-L1 expression ≥ 1% populations.

Unless otherwise specified, the demography, protocol deviations, baseline characteristics, efficacy [redacted] will be analyzed in the randomized subjects and in the randomized subjects with PD-L1 expression ≥ 1% populations.

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, discrete variables will be tabulated by the frequency and proportion of subjects falling into each category, grouped by treatment. Percentages given in these tables will be rounded to the first decimal and, therefore, may not always sum to 100%. Percentages less than 0.1 will be indicated as '< 0.1'. Continuous variables will be summarized by treatment group using the mean, standard deviation, median, minimum, and maximum values.

Time-to-event variables (e.g. time-to resolution) will be analyzed using the Kaplan-Meier technique. When specified, the median will be reported along with 95% CI using Brookmeyer and Crowley method[13] (using log-log transformation for constructing the confidence intervals[14]).

The conventions to be used for imputing missing and partial dates for analyses requiring dates are described in Section 8.

7.1.1 Adverse Events, Serious Adverse Events, Multiple events, Select Adverse Events, Other Events of Special Interest and Immune-Mediated Adverse Events

Drug-related AEs are those events with relationship to study drug "Related", as recorded on the CRF. If the relationship to study drug is missing, the AE will be considered as drug-related.

Serious adverse events consist of AEs deemed serious by the Investigator and flagged accordingly in the CRF and clinical database.

Adverse events leading to study drug discontinuation are AEs with action taken regarding study drug(s) = "Drug was discontinued".

Adverse events leading to dose delay are AEs with action taken regarding study drug(s) = "Drug was delayed".

Adverse events leading to dose reduction are AEs with action taken regarding study drug(s) = "Dose was reduced".

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), and the most recent version of the dictionary at the time of the database lock will be used. Adverse events results will be graded for severity using NCI Common Terminology Criteria for Adverse Events (CTCAE) and the most recent version of the criteria at the time of the database lock will be used.

In the AE summary tables, unless otherwise specified, subjects will be counted only once at the Preferred Term (PT), only once at the System Organ Class (SOC), and only once at subject level for the counting of total number of subjects with an AE. The AE tables will be sorted by the SOCs and then PTs. SOC will be ordered by descending frequency overall and then alphabetically. PTs will be ordered within SOC by descending frequency overall and then alphabetically. The sorting will be done based on the 'Any Grade' column of the experimental arm when arms are presented side-by-side.

Unless otherwise specified, the AE summary tables will be restricted to on-treatment events regardless of the causality.

Analyses that take into account the multiple occurrences of a given adverse event will be conducted (see Section 7.6.9). To prepare these analyses, the CRF data will be processed according to standard BMS algorithms[15] in order to collapse adverse event records into unique records based on the preferred term. These data will be presented as the rate per 100 person-years of exposure. These analyses will take into account all on-treatment events (allowing more than 1 event per subject) and the total exposure time. The person-year exposure will be computed as the sum over the subjects' exposure expressed in years where the exposure time is defined as

- (Date of last dose of study treatment - date of first dose of study treatment + 31 days (or 101 days, depending on the analysis))/365.25, for subject who are off study treatment and were followed for at least 30 days (or 100 days, depending on the analysis) after last dose of study treatment.
- (Last known alive date - date of first dose of study treatment +1)/365.25, for subjects who are still on-treatment or who are off study treatment and were followed less than 30 days (or 100 days depending on the analysis) after last dose of study treatment.

7.1.1.1 Select Adverse Events

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (e.g. pulmonary events, gastrointestinal events categories, etc.). AEs that may differ from or be more severe than AEs caused by non-immunotherapies and AEs whose early recognition and management may mitigate severe toxicity are included as select AEs. Categories of select AEs may include subcategories (e.g. adrenal disorders, diabetes, pituitary disorders, and thyroid disorders are subcategories of the endocrine event category).

The list of MedDRA preferred terms used to identify select adverse events is revisited quarterly and updated accordingly. The preferred terms used for the selection at the time of the database lock will be provided by categories/subcategories.

In addition to the frequency and worst severity of select AEs, time-to onset, time-to resolution, and time-to resolution where immune modulating medication was initiated will be analyzed for each specific category/subcategory of drug-related select AEs when applicable.

Further details on the definitions time-to onset and time-to resolution are described in 0.

[redacted]

7.1.2 Laboratory Tests

Clinical laboratory parameters (hematology, serum chemistry and electrolytes) will be evaluated.

Laboratory tests will be graded using the NCI Common Terminology Criteria, version 4.0.

Clinical laboratory data will be first analyzed using International System of Units (SI). Analyses will be repeated using US conventional units.

In the laboratory summary tables, unless otherwise specified, subjects will be counted only once for each lab parameter according to their worst on treatment CTC grade (worst being the highest CTC grade). The laboratory tables and listings will be sorted by laboratory category, laboratory subcategory and laboratory test code sequence number.

[redacted]

7.2 Study Conduct

7.2.1 Accrual

Enrollment by country and site, and enrollment by month will be summarized and listed for all enrolled subjects.

A by-subject listing of batch numbers for all treated subjects will be provided.

7.2.2 Relevant Protocol Deviations

The following programmable deviations from inclusion and exclusion criteria will be considered as relevant protocol deviations.

The Relevant Protocol Deviations will be summarized and listed for all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1%, by treatment group and overall.

At Entrance:

- Subjects with presence of disease at baseline
- [redacted]
- Subject randomized more than 120 days after IUC surgery
- Subject with ineligible pathologic stage at resection

On-study:

- Subjects receiving anti-cancer therapy* (chemotherapy, hormonal therapy, immunotherapy, standard or investigational agents for treatment of cancer) while on study therapy
- Subjects treated differently than as randomized (subjects who received the wrong treatment, excluding the never treated)

* except single dose of BCG or other chemotherapy given via intravesical route.

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.

7.3 Study Population

Analyses in this section will be tabulated for all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1% by treatment group as randomized, unless otherwise specified.

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. This analysis will be performed on the all enrolled subjects population only pooled together.

Number of subjects randomized but not treated along with the reason for not being treated will be tabulated by treatment group as randomized. This analysis will be performed only on the all treated subjects population and on the all treated subjects with PD-L1 expression ≥ 1%.

Number of subjects who discontinued study treatment along with corresponding reason will be tabulated by treatment group as treated. Reason for discontinuation will be derived from subject status CRF page. This analysis will be performed only on the all treated subjects population and on the all treated subjects with PD-L1 expression ≥ 1%.

A by-subject listing for all enrolled subjects will also be provided, showing whether the subject was randomized/treated along with the reason for not being randomized/treated.

A by-subject listing for all treated subjects will be provided showing the subject's off treatment date and whether the subject continue in the treatment period/study along with the reason for going off treatment period/study.

7.3.2 Demographics and Other Baseline Disease Characteristics

The following demographic and baseline disease characteristics will be summarized and listed for all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1% by treatment group as randomized:

- Age (continuous, in years)
- Age categorization (< 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85, ≥ 75, ≥ 65 years)

- Sex (Male vs. Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Country by geographic region
- [redacted]
- Baseline weight (continuous, in kg)
- Smoking status (former/current smoker, never smoker, unknown, not reported)
- Tumor type (urinary bladder, renal pelvis, ureter)
- **Minor histological variants (none, adenocarcinoma, squamous cell carcinoma, small cell carcinoma, miropapillary, nested, plasmacytoid, sarcomatoid, other)**
- Time from initial disease diagnosis to randomization (< 1 year, ≥ 1 year)
- **Pathologic stage (tumors, nodes and node density) at resection (from CRF database)**
- Baseline hemoglobin (< LLN, ≥ LLN)
- Baseline Creatinine Clearance (CrCL)* (< 30, ≥ 30 and < 60, ≥ 60ml/min)
- Baseline PD-L1 expression **status (< 1%, ≥ 1% and < 5%, ≥ 5% and < 10%, ≥ 10%, ≥ 5%, ≥ 1%, indeterminate, not evaluable) (from clinical database)**

* The creatinine clearance will be calculated using Cockroft-Gault formula, defined as:

$$\mathrm{CrCL(ml/min)} = \frac{(140 - \mathrm{age\,(in\ years)}) * \mathrm{weight\,(in\ kg)}}{72 * \mathrm{serum\ creatinine\,(in\ mg/dL)}}$$

for males and

$$\mathrm{CrCL(ml/min)} = \frac{(140 - \mathrm{age\,(in\ years)}) * \mathrm{weight\,(in\ kg)}}{72 * \mathrm{serum\ creatinine(in\ mg/dL)}} * 0.85$$

for females. Baseline weight will be used.

Summary tables (cross-tabulation) by treatment group as randomized for stratification factor, as given here below, will be provided to show any discrepancies between what was reported through IRT vs. other data sources at baseline. This summary will be performed based on all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1%.

- PD-L1 expression level (≥ 1% vs. < 1%/indeterminate)
- [redacted]

A listing of randomization scheme presenting randomized treatment group and as treated treatment group will be provided for all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1%.

7.3.3 Medical History

A by-subject listing of general medical history will be provided for all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1%.

7.3.4 Prior Therapy Agents

Prior cancer therapy will be summarized by treatment group and overall for all randomized subjects as well as on the all randomized subjects with PD-L1 expression ≥ 1%.

The following will be summarized by treatment group for randomized subjects:

- Prior agent received and by setting
- Prior regimen received and by setting
- Prior Cisplatin therapy (CRF database)
- Reason not treated with Prior Cisplatin
- IUC surgery summary (No Surgery vs. Radical Cystectomy vs. Radical Cystoprostatectomy vs. Radical Nephroureterectomy vs. Radical Ureterectomy vs. Other)
- Time from IUC surgery to randomization (0 - 30, > 30 - 60, > 60 - 90, > 90 - 120, > 120 days)
- Prior radiotherapy (Yes or No)

Prior systemic cancer therapy will be summarized by treatment group and overall and listed by subject.

Prior radiotherapy and prior surgery related to cancer will be listed by subject.

7.3.5 Physical Examinations

Subjects with abnormal baseline physical examination will be listed by subject.

7.3.6 Baseline Physical Measurements

Baseline physical measurements will be listed by subject.

7.4 Extent of Exposure

Analyses will be performed by treatment group as treated in all treated subjects as well as in all treated subjects with PD-L1 expression ≥ 1% , unless otherwise specified.

Listings will include all available exposure data.

7.4.1 Administration of Study Therapy

Study therapy definition is given in Table 7.4.1-1.

Table 7.4.1-1: Administration of Study Therapy

	Nivolumab / Placebo
Dosing schedule per protocol	240mg every 2 weeks
Dose	Dose (mg) is defined as the ratio of Total Volume Infused with Total Volume Prepared x 240 in mg. Volume infused/prepared in mL at each dosing date is collected on the CRF.
Cumulative Dose	Cum dose (mg) is the sum of the doses (mg) administered to a subject during the treatment period.
Relative dose intensity (%)	[Cum dose (mg)/((Last dose date - Start dose date + 14) x 240 / 14)] x 100
Duration of study therapy	Last dose date - Start dose date +1

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

- **Time from randomization to first dose of study therapy (0 to 3 days, > 3 to 7, > 7 to 14, > 14 to 21, > 21 to 28, > 28)**
- Number of doses received
- Cumulative dose
- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%
- Duration of study therapy

Time to treatment discontinuation will be summarized and presented by treatment group as treated using a Kaplan-Meier curve whereby the last dose date will be the event date for subjects who discontinued study therapy. Subjects who are still on study therapy will be censored on their last dose date. Median duration of study therapy and associated 95% CI will be provided.

A by-subject listing of dosing of study medication (record of study medication, infusion details, dose changes **and reason for discontinuation**) will be also provided.

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delays

Each nivolumab infusion may be delayed. Treatment may be delayed for up to a maximum of 6 weeks from the last dose. 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 nivolumab.

The following parameters will be summarized by treatment group:

- Number of subjects with at least one dose delayed,
- Number of dose delays per subject,
- Reason[3)a)] for dose delay,
- Length[b)] of dose delay.

a) Reason for dose delay will be retrieved from CRF dosing pages.

b) Length of dose delay is defined as (duration of previous cycle in days - 14). Dose delays will be divided into following categories: 4 - < 8 days, 8 - < 15 days, 15 - < 42, ≥ 42 days.

7.4.2.2 Infusion Interruptions and Rate Changes

Each nivolumab 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 group:

- Number of subjects with at least one dose infusion interruption, the reason for interruption, and the number of infusion interruptions per subject.
- Number of subjects with at least one IV infusion rate reduction, the reason for reduction and the number of infusion with IV rate reduction per subject.

7.4.2.3 Dose Escalations

Dose escalations (within subject) are not permitted for nivolumab.

7.4.2.4 Dose Reductions

Dose reductions (within subject) are not permitted for nivolumab.

7.4.2.5 Dose Omissions

Dose omissions (within subject) are not permitted for nivolumab.

[redacted]

[redacted]

7.4.3.2 Subsequent Cancer Therapy

The following summary table will be provided overall and by treatment group as randomized in all randomized subjects as well as all randomized subjects with PD-L1 expression ≥1%:

- Number and percentage of subjects receiving subsequent cancer therapies including the following categories:
  - Number and percentage of subjects receiving subsequent systemic therapy, overall, by type of therapy and by drug name:
    - Number and percentage of subjects receiving subsequent immunotherapy (anti-PD1 agents, anti-PD-L1 agents, anti-CTLA-4 agents and others), overall and by drug name
    - Number and percentage of subjects receiving subsequent BCG or other intravesical chemotherapy (approved and investigational), overall and by drug name
    - Number and percentage of subjects receiving subsequent other agents (approved and investigational), overall and by drug name
  - Number and percentage of subjects receiving subsequent surgery for treatment of tumors, overall, by type of surgery
  - Number and percentage of subjects receiving subsequent radiotherapy for treatment of tumors
  - Number and percentage of subjects receiving any combination of the above

A by-subject listing of subsequent cancer therapy will also be produced.

7.5 Efficacy

Analyses in this section will be tabulated by treatment group as randomized, unless otherwise specified.

Unless otherwise specified, efficacy analyses will be performed independently for:

- all randomized subjects
- all randomized subjects with PD-L1 expression ≥1%.

Unless stated otherwise, whenever a stratified analysis is specified, the following stratifications factors (recorded at randomization as per IRT) will be used:

- Receipt of neo-adjuvant cisplatin based chemotherapy for IUC (Yes vs. No)
- Pathologic status of disease in lymph nodes (N+ vs. N0/x with < 10 Nodes Removed vs. N0 with ≥ 10 Nodes Removed)

Only for analysis in all randomized subjects:

- PD-L1 expression level (≥ 1% vs. < 1%/indeterminate)

For assessing the first secondary objective (with overall survival as secondary endpoint) of this study, a hierarchical testing procedure[16] will be used so that the overall experiment-wise Type I error rate is two-sided 0.05.

Confidence intervals (CI) for primary and secondary endpoints analyses included in hierarchy will be based on nominal significance level adjusted for primary endpoints and interim analyses to preserve overall type one error rate.

Alpha (α) for the CI will be the same as nominal significance level for hypothesis testing. CIs for other endpoints will be at the two-sided 95% level. All p-values reported will be two-sided. P-values will be rounded to the fourth decimal place. Point estimates and confidence bounds for efficacy variables will be rounded to the second decimal place.

7.5.1 Disease-Free Survival

The primary objective of the study is to compare the disease-free survival between treatment groups in each co-primary populations (all randomized subjects and all randomized subjects with PD-L1 expression ≥ 1%).

7.5.1.1 Analysis of Disease-Free Survival

In case of formal comparison (interim or final), DFS will be compared between the treatment groups via stratified log-rank test in the appropriate co-primary population at an overall two-sided $\alpha = 0.025$ level for each co-primary population. The stratification factors are summarized in the section 7.5.

An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. The estimate of the DFS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. Ties will be handled using the exact method. A two-sided (1-adjusted α)% CI (adjusted for interim) for the hazard ratio will also be presented.

The DFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median DFS in each treatment group will be computed via the log-log transformation method. DFS rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood[17] formula for variance derivation and on log-log transformation applied on the survivor function[18].

The primary definition of DFS adjusting for subsequent anticancer therapy will be used in this analysis. The two-sided log-rank p-value will be reported.

The source of DFS event (recurrence or death) will be summarized by treatment group. The status of subjects who are censored (as per primary definition of DFS) in the DFS KM analysis will be tabulated for each treatment group including the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)
- No baseline disease assessment
- No on-study disease assessment and no death
- Received subsequent anticancer therapy
- New non-urothelial carcinoma primary cancer

A by-subject listing will be presented including treatment group, DFS duration under the primary definition, whether the subject was censored under the primary definition, and if censored, the reason.

7.5.1.2 Supportive Analyses of Disease-Free Survival

The following sensitivity analyses will be conducted using the primary and secondary definition of DFS:

1) DFS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.

2) DFS using an unstratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.

3) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization, which, by definition, will be balanced across treatment groups, will still be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include:
   - Age (< 65 vs. ≥ 65 year)
   - Gender (Male vs. Female)
   - Baseline ECOG status (0 vs. 1 vs. 2)
   - Pathological status (pT0-2 vs. pT3 vs. pT4)

   The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 95% CIs.

4) DFS for subjects with no relevant protocol deviations. This analysis will be conducted only if there are more than 10% subjects with relevant protocol deviations. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.

5) DFS accounting for missing disease assessments prior to DFS event (recurrence or death). This analysis will be performed only if at least 10% of DFS events (in the population of interest) have two or more missing prior disease assessment (no protocol specified disease assessment imaging completed at the expected timepoint). It will apply the following restriction to the primary definition in case a subject has two or more missing disease assessments, the subject will be censored at the last disease assessment date prior to DFS event.

6) Delayed effect of immunotherapy interventions may cause a late separation in the DFS KM curves and non-proportional hazards. DFS will be compared between treatment groups via two-sided 0.05 stratified weighted log-rank test among subjects. The two-sided stratified weighted log-rank p-value will be reported using G (rho = 0, gamma = 1) weights, in the terminology of Fleming and Harrington[19].

   The Fleming Harrington test can be unstable, so it is possible, though uncommon, that the p-value for this trial will not be estimable.

   The estimate of the DFS hazard ratio in the period before and following 6 months will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. In this model, period is a binary variable indicating pre- vs. post-6 months. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio will also be presented.

7) 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 non-constant treatment effect. In that case, additional exploratory analyses may be performed.

7.5.1.3 Subset Analyses of Disease-Free Survival

The influence of baseline and demographic characteristics on the treatment effect will be explored via exploratory subset analyses. The median DFS based on KM product-limit method along with two-sided 95% CIs will be produced for the following subgroups:

- PD-L1 Status (Positive ≥ 1% vs. Negative < 1% vs. Indeterminate / Not Evaluable; source: clinical database)
- Use of prior neo-adjuvant cisplatin therapy (Yes vs. No; source: CRF)
- Prior Cisplatin status (received Neo-Adjuvant, Unwilling vs. Ineligible vs. Other; source: CRF)
- Use of any prior neo-adjuvant systemic therapy (Yes and No) (source: CRF)
- Pathological lymph node status (N+ vs. N0/x with < 10 Nodes Removed vs. N0 with ≥ 10 Nodes Removed; source: CRF)
- Pathological status (pT0-2 vs. pT3 vs. pT4; source:CRF)
- Age (< 65, ≥ 65 and < 75, ≥ 75 years)
- Region (US, EU, ROW).

- Gender (Male, Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Baseline ECOG status (0, 1, 2)
- Smoking Status (Never, Former/Current, Unknown)
- Baseline Hemoglobin (< LLN vs. ≥ LLN)
- Baseline Creatinine Clearance (≥ 60 vs. < 60ml/min)
- Time from IUC surgery to randomization (0 - 30, > 30 - 60, > 60 - 90, > 90 - 120, > 120 days)
- Initial tumor origin (Urinary Bladder vs. Renal Pelvis, Ureter)
- Minor histological variants (Presence vs. Absence)

A forest plot of the DFS hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above. The analysis comparing treatment (i.e., Hazard Ratio) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.1.4 Current Status of DFS Follow-up

The currentness of follow-up for DFS, defined as the time between last evaluable disease assessment (i.e., regardless of initiation of subsequent therapy or new non-urothelial carcinoma primary cancer) and cutoff date (defined by last patient last visit date), will be summarized by treatment group in months. Subjects who had DFS events and subjects with last evaluable disease assessment date on or after data cut-off date will have zero value for currentness of follow-up.

A by-subject listing will also be produced to accompany the subject time from last evaluable disease assessment.

7.5.2 Overall Survival

The key secondary objective of the study is to compare the overall survival between treatment groups in each co-primary populations (all randomized subjects and all randomized subjects with PD-L1 expression ≥ 1%).

7.5.2.1 Analysis of Overall Survival

Overall survival will be compared between the treatment groups at the interims and final analyses, using stratified log-rank test, as defined in Section 5. The stratification factors will be those used in the analysis of DFS.

An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. The estimate of the OS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. Ties will be handled using the exact method. A two-sided (1-adjusted α)% CI (adjusted for interims) for the hazard ratio will also be presented.

The OS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median OS in each treatment group

will be computed via the log-log transformation method. OS rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula17 for variance derivation and on log-log transformation applied on the survivor function[18].

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

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)

A by-subject listing will be presented including treatment group, first and last dose date, whether the subject died, and if censored, the reason, event/censored date and OS duration.

7.5.2.2 Supportive Analyses of Overall Survival

The following sensitivity analyses will be conducted using the primary definition of OS:

1) OS using stratification factors as obtained from the baseline CRF pages (instead of IRT). The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IRT) and baseline are not concordant for at least 10% of the randomized subjects.

2) OS using an unstratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.

3) A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization, which, by definition, will be balanced across treatment groups, will still be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include:
   - Age (< 65 vs. ≥ 65 year)
   - Gender (Male vs. Female)
   - Baseline ECOG status (0 vs. 1 vs. 2)
   - Pathological status (pT0-2 vs. pT3 vs. pT4)

   The level of the covariate normally associated with the worst prognosis will be coded as the reference level. The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 95% CIs.

4) OS for subjects with no relevant protocol deviations. This analysis will be conducted only if there are more than 10% subjects with relevant protocol deviations. The hazard ratio associated with treatment will be presented along with the associated two-sided 95% CIs.

5) 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 non-constant treatment effect. In that case, additional exploratory analyses may be performed.

7.5.2.3 Subset Analyses of Overall Survival

The influence of baseline and demographic characteristics on the treatment effect will be explored via exploratory subset analyses. The median OS based on KM product-limit method along with two-sided 95% CIs will be produced for the same subgroups used for DFS (see Section 7.5.1.2).

A forest plot of the OS hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above. The analysis comparing treatment (i.e., Hazard Ratio) will be conducted if the number of subjects in the subgroup category is more than 10.

7.5.2.4 Current Status of OS Follow-up

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

The currentness of follow-up for survival, defined as the time between last OS contact (i.e., last known alive date or death date) and cutoff date (defined by last patient last visit date), will be summarized in months. Subjects who died and subjects with last known alive date on or after data cut-off date will have zero value for currentness of follow-up.

Minimum follow-up of OS, defined as the time from cutoff date to last subject's randomization date, will be summarized in months.

7.5.3 Interim/Final Analyses of Disease-Free Survival and Overall Survival

In addition to the formal planned interim and final analyses of DFS and OS, the Data Monitoring Committee (DMC) will have access to periodic un-blinded interim reports of efficacy and safety to allow a risk/benefit assessment. Details are included in the DMC charter.

The DMC will review the safety and efficacy data from the informal interim analyses and will determine if the study should continue with or without changes or if accrual should be stopped. Subject enrollment will continue while waiting for the DMC's decisions.

The chair of the DMC and the sponsor can call an unscheduled review of the safety data.

At the time of the formal interim analysis for superiority of DFS, the DMC may recommend continuing or stopping the trial.

- If the trial continues beyond the formal interim analysis, the nominal critical point for the final DFS analysis will be determined using the recalculated information fraction at the time of the interim analysis, as described above. The final DFS hazard ratio and corresponding confidence interval will be reported whereby the confidence interval will be adjusted accordingly (i.e. using the recalculated nominal αlevel at the final analysis).

- If the trial is stopped for superiority of DFS at the interim, the p-value from the interim stratified log-rank test will be considered the final primary analysis result.

  If allowed per hierarchical testing procedure, at the time of the formal interim analyses for superiority of OS, the DMC may recommend continuing or stopping the trial (see Section 5.1).
- If the trial continues beyond the two formal interim analysis, the nominal critical point for the final OS analysis will be determined using the recalculated information fraction at the time of each interim analyses, as described above. The final OS hazard ratio and corresponding confidence interval will be reported whereby the confidence interval will be adjusted accordingly (i.e. using the recalculated nominal αlevel at the final analysis).
- If the trial is stopped for superiority of OS at one of the interim analyses, the p-value from the interim stratified log-rank test will be considered the final primary analysis result.

The same above procedure applies independently in each co-primary population.

7.5.4 Non-Urothelial Tract Recurrence Free-Survival

Similar analysis to primary endpoint of DFS will be conducted for NUTRFS, but no formal comparison between treatment groups will be done.

The NUTRFS function for each treatment group will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median NUTRFS in each treatment group will be computed via the log-log transformation method. NUTRFS rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula[17] for variance derivation and on log-log transformation applied on the survivor function[18].

The primary definition of NUTRFS adjusting for subsequent anticancer therapy will be used in this analysis. The two-sided log-rank p-value will be reported.

The source of NUTRFS event (recurrence or death) will be summarized by treatment group. The status of subjects who are censored (as per primary definition of NUTRFS) in the NUTRFS KM analysis will be tabulated for each treatment group including the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)
- No baseline disease assessment
- No on-study disease assessment and no death
- Received subsequent anticancer therapy
- New non-urothelial carcinoma primary cancer

7.5.5 Disease-Specific Survival

No formal comparison between treatment groups will be done.

Given the presence of competing risks, cause-specific HRs and corresponding two-sided 95% CIs will be estimated using a Cox proportional hazards model, with treatment group as a single covariate, stratified by same stratification factors as used for DFS and OS.

The cumulative incidence of DSS for each treatment group will be displayed. Rates at fixed time points (e.g. 6 months, depending on the minimum follow-up) will be derived from the cumulative incidence curve estimate and corresponding confidence interval will be derived.

For completeness, the cumulative incidence of death non due to disease will be also displayed for each treatment group.

The source of DSS event (death due to disease) will be summarized by treatment group. The status of subjects who are censored in the DSS cause-specific hazard analysis will be tabulated by treatment group using following categories:

- Death due to other cause
- On-study (on treatment, in follow-up)
- Off-study: (lost to follow-up, withdrew consent).

7.6 Safety

Analyses in this section will be tabulated for all treated subjects as well as for all treated subjects with PD-L1 expression ≥ 1% by treatment group as treated, unless otherwise specified.

7.6.1 Deaths

Deaths will be summarized by treatment group:

- All deaths, reasons for death.
- Deaths within 30 days of last dose received, reasons for death.
- Deaths within 100 days of last dose received, reasons for death.

A by-subject listing of deaths will be provided for the all enrolled subjects population.

7.6.2 Serious Adverse Events

Serious adverse events will be summarized by treatment group:

- Overall summary of SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

All analyses will be conducted using the 30-day safety window.
A by-subject SAE listing will be provided for the all enrolled subjects population.

7.6.3 Adverse Events Leading to Discontinuation of Study Therapy

AEs leading to discontinuation will be summarized by treatment group:

- Overall summary of AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

The analyses will be conducted using the 30-day safety window.

A by-subject AEs leading to discontinuation listing will be provided.

7.6.4 Adverse Events Leading to Dose Modification

AEs leading to dose delay/reduction will be summarized by treatment group:

- Overall summary of AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of related AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

The analysis will be conducted using the 30-day safety window.

A by-subject AEs leading to dose delay/reduction listing will be provided.

[redacted]

7.6.15 Pregnancy

A by-subject listing of pregnancy tests results will be provided for randomized female subjects.

7.6.16 Adverse Events by Subgroup

Overall summary of any AEs and drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT and for each treatment group for the following subgroups:

- Sex (Male vs. Female)
- Race
- Age (< 65 vs. 65 - < 75 vs. 75 - < 85 vs. ≥ 85 vs. ≥ 75 vs. ≥ 65)
- Region (US, EU, ROW)

These analyses will be conducted using the 30-day safety window only.

[redacted]

[redacted]

7.8.1 Distribution of PD-L1 Expression

Descriptive statistics of PD-L1 expression will be performed in all randomized subjects, unless otherwise specified:

- Listing of all PD-L1 IHC data.
- Summary of tumor specimen acquisition and characteristics.
- Summary statistics of PD-L1 expression in all randomized subjects with quantifiable PD-L1 expression.
- Cumulative distribution plot of baseline PD-L1 expression versus population percentile in all randomized subjects with quantifiable PD-L1 expression.
- Waterfall plot of Individual PD-L1 expression in all randomized subjects with quantifiable PD-L1 expression.

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[21]
- For missing and partial adverse event resolution dates, imputation will be performed as follows (these conventions may change):
  - If only the day of the month is missing, the last day of the month will be used to replace the missing day. If the imputed date is after the death date or the last known alive date, then the latest known alive date or death date is considered as the resolution date.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- Missing and partial non-study medication domain dates will be imputed using the derivation algorithm described in Section 4.1.3 of BMS Non-Study Medication Domain Requirements Specification[22].
- Missing and partial radiotherapy and surgery dates will be imputed using algorithm described in APPENDIX 2.
- 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 alive date 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 alive date.
  - 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 recurrence after start of study therapy, 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. In case of the date of death is present and complete, the imputed recurrence date will be compared to the date of death. The minimum of the imputed recurrence date and date of death will be considered as the date of recurrence.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- For date of recurrence to prior therapies, 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.
- For other partial/missing dates, the following conventions were 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-to onset, time-to resolution) will be calculated as follows:

Duration = (Last date - first date + 1)

Last known alive date will be defined based on all appropriate dates collected on the CRF.

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

The analyses described in this SAP are potential analyses to be included in study-specific CSR.

Refer to the Data Presentation Plan for mock-ups of all tables and listings.