

I4T-MC-JVDC Statistical Analysis Plan Version 3

A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of Ramucirumab plus Docetaxel Versus Placebo plus Docetaxel in Patients with Locally Advanced or Unresectable or Metastatic Urothelial Carcinoma Who Progressed on or After Platinum-Based Therapy

NCT02426125

Approval Date: 13-Feb-2018

**1. Statistical Analysis Plan for Clinical Study:  
I4T-MC-JVDC: A Phase 3, Randomized, Double-Blind,  
Placebo-Controlled Study of Ramucirumab plus  
Docetaxel Versus Placebo plus Docetaxel in Patients with  
Locally Advanced or Unresectable or Metastatic  
Urothelial Carcinoma Who Progressed on or After  
Platinum-Based Therapy**

**Confidential Information**

The information contained in this document is confidential and the information contained within it may not be reproduced or otherwise disseminated without the approval of Eli Lilly and Company or its subsidiaries.

**Note to Regulatory Authorities:** this document may contain protected personal data and/or commercially confidential information exempt from public disclosure. Eli Lilly and Company requests consultation regarding release/redaction prior to any public release. In the United States, this document is subject to Freedom of Information Act (FOIA) Exemption 4 and may not be reproduced or otherwise disseminated without the written approval of Eli Lilly and Company or its subsidiaries.

**Ramucirumab (LY3009806) Urothelial Carcinoma**

Phase 3, randomized, double-blind, placebo-controlled study of ramucirumab plus docetaxel versus placebo plus docetaxel in patients with locally advanced or unresectable or metastatic urothelial carcinoma who progressed on or after one prior first-line platinum-based chemotherapy. Patients will be randomized 1:1 to receive one of these study regimens on Day 1 of each 21-day cycle: ramucirumab (10 mg/kg) I.V. plus docetaxel (75 mg/m<sup>2</sup>) I.V. OR placebo (10 mg/kg volume equivalent) I.V. plus docetaxel (75 mg/m<sup>2</sup>) I.V. Provided no prespecified discontinuation criteria have been met (including radiographic documentation of disease progression, toxicity requiring cessation, protocol noncompliance, or withdrawal of consent), treatment with docetaxel may continue for up to six 21-day cycles; further cycles of docetaxel (up to 4 additional cycles [maximum of 10 cycles total]) may be administered with approval of the Lilly clinical research physician/clinical research scientist or designee. Treatment with ramucirumab or placebo (monotherapy) may continue on 21-day cycles until at least one discontinuation criterion is met.

Eli Lilly and Company  
Indianapolis, Indiana USA 46285  
Protocol I4T-MC-JVDC  
Phase 3

Statistical Analysis Plan electronically signed and approved by Lilly: 27 January 2016

Statistical Analysis Plan Version 2 electronically signed and approved by Lilly:  
11 May 2017

Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on date  
provided below.

Approval Date: 13-Feb-2018 GMT

## 2. Table of Contents

Section	Page
1. Statistical Analysis Plan for Clinical Study: I4T-MC-JVDC: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of Ramucirumab plus Docetaxel Versus Placebo plus Docetaxel in Patients with Locally Advanced or Unresectable or Metastatic Urothelial Carcinoma Who Progressed on or After Platinum-Based Therapy.....	1
2. Table of Contents.....	2
3. Revision History.....	6
4. Study Objectives.....	7
4.1. Primary Objective.....	7
4.2. Secondary Objectives.....	7
4.3. Exploratory Objectives.....	7
5. A Priori Statistical Methods.....	8
5.1. Determination of Sample Size.....	8
5.2. General Considerations.....	10
5.2.1. Definitions of Analysis Variables.....	11
5.2.1.1. Efficacy Analysis Variables.....	12
5.2.1.2. Safety Analysis Variables.....	14
5.2.1.3. Patient-Reported Outcome Analysis Variables.....	16
5.3. Adjustments for Covariates.....	17
5.4. Handling of Dropouts or Missing Data.....	18
5.5. Multicenter Studies.....	20
5.6. Multiple Comparisons/Multiplicity.....	20
5.7. Study Patients.....	20
5.7.1. Analysis Populations.....	20
5.7.2. Important Protocol Deviations.....	21
5.8. Demographic and Other Baseline Characteristics.....	22
5.9. Concomitant Medications.....	23
5.10. Treatment Compliance.....	23
5.11. Efficacy Analyses.....	23
5.11.1. Primary Efficacy Analyses.....	23
5.11.2. Secondary Efficacy Analyses.....	25
5.11.2.1. Supportive Analyses of Primary Efficacy Endpoint.....	25
5.11.2.2. Key Secondary Efficacy Analyses.....	25
5.11.2.3. Analyses of Other Efficacy Endpoints.....	26

5.11.3.	Additional Overall Survival Analyses .....	26
5.11.4.	Subgroup Analyses .....	28
5.12.	Post-Discontinuation Therapy .....	28
5.13.	Patient-reported Outcome Analyses.....	28
5.13.1.	EORTC QLQ-C30 .....	28
5.13.2.	EQ-5D-5L.....	29
5.14.	Safety Evaluation .....	29
5.14.1.	Exposure.....	29
5.14.2.	Adverse Events .....	29
5.14.2.1.	Overall Summary of Adverse Events .....	29
5.14.2.2.	Treatment-Emergent Adverse Events.....	30
5.14.3.	Deaths, Serious Adverse Events, and Other Significant Adverse Events .....	30
5.14.4.	Clinical Laboratory Evaluation.....	31
5.14.5.	Hospitalizations and Transfusions .....	31
5.14.6.	Vital Signs, Physical Findings, and Other Observations Related to Safety .....	31
5.14.7.	Subgroup Analyses .....	31
5.15.	Pharmacokinetics and Immunogenicity .....	32
5.15.1.	Pharmacokinetics .....	32
5.15.2.	Immunogenicity .....	32
5.16.	Translational Research .....	32
5.17.	Interim Safety Analysis .....	33
5.18.	Clinical Trial Registry Analyses.....	34
6.	Unblinding Plan .....	35
7.	References .....	37

**Table of Contents**

<b>Table</b>		<b>Page</b>
Table JVDC.5.1.	Study Design and Operating Characteristics.....	10
Table JVDC.5.2.	Censoring Rule of PFS Primary Analysis.....	12
Table JVDC.5.3.	Censoring Rules for PFS Sensitivity Analysis Definitions.....	13
Table JVDC.5.4.	Analysis Populations.....	21

**Table of Contents**

<b>Appendix</b>		<b>Page</b>
Appendix 1.	Flow Chart of PFS Censoring Rules .....	38
Appendix 2.	Technical Details on Independent Increment of Survival Times .....	39

### 3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved prior to the first unblinding, to allow execution of activities related to the interim analysis.

Version 2 was approved prior to the primary progression-free survival (PFS) analysis to update the components listed below:

- Added analysis for independent central review of radiological data (referred to as independent assessment) for all tumor-related events since decision was made to conduct a full central read for this study
- Added definition of time to progression since it was omitted in previous version
- Removed/revised subgroups used for adjustments for covariates and subgroup analysis of efficacy due to the changing landscape of treatment options
- Modified the restricted mean difference analysis to provide more details on the planned analysis
- Clarified REGION breakdown so it would be clearer which countries fell in each region
- Clarified and added/removed/revised the variables and analysis of Patient-Reported Outcomes since was not available previously
- Modified analysis population table to made it clear which population would be used
- Added the definition and analysis of important protocol deviations after changes were made prior to study team unblinding
- Modified Per Protocol Population definition based on revised definitions of important protocol deviations
- Modified time to deterioration analysis for European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire - 30 items (EORTC QLQ-C30)
- Added references

Version 3 was approved prior to the primary overall survival analysis. The primary PFS lock occurred on 19 May 2017 and interim overall survival was assessed per design.

After discussions with the Food and Drug Administration, the SAP is being revised to add additional overall survival analyses. See Section 5.11.3 (Additional Overall Survival Analyses) for the details of these analyses. [Appendix 2](#) was added to give the technical details of these new analyses.

In Section 5.7.2, one of the important protocol violations was revised to match correctly with the violation of Inclusion Criterion 2. This had been written incorrectly in the previous version of the SAP.

## 4. Study Objectives

### 4.1. Primary Objective

The primary objective of this study is to compare the PFS of ramucirumab in combination with docetaxel with the PFS of placebo in combination with docetaxel, in patients with locally advanced or unresectable or metastatic urothelial carcinoma who have had disease progression on or after one prior first-line platinum-based chemotherapy.

### 4.2. Secondary Objectives

The secondary objectives of this study per protocol are to compare each of the following variables between the treatment arms:

- overall survival (OS) time
- objective response rate (ORR; complete response [CR] + partial response [PR]) and disease control rate (DCR)
- duration of response (DOR)
- safety profile
- patient-reported outcome (PRO) measures (EORTC QLQ-C30 and EuroQol 5 Dimension, 5 Level [EQ-5D-5L] questionnaires)

Secondary objectives also include the evaluation of:

- the pharmacokinetic (PK) profile of ramucirumab
- the immunogenicity of ramucirumab (anti-ramucirumab antibodies)

### 4.3. Exploratory Objectives

The exploratory objectives of this study are to:

- assess the change in tumor size in patients with measurable disease
- examine biomarkers relevant to ramucirumab, angiogenesis, and the disease state, and to correlate these markers to clinical outcome
- time to progression
- the PK of docetaxel (following dosing at 60 mg/m<sup>2</sup> or 75 mg/m<sup>2</sup>)



## 5. A Priori Statistical Methods

### 5.1. Determination of Sample Size

The primary objective of this study is to compare the PFS of ramucirumab in combination with docetaxel with the PFS of placebo in combination with docetaxel, in patients with locally advanced or unresectable or metastatic urothelial carcinoma who have had disease progression on or after 1 prior first-line platinum-based chemotherapy.

The study will include 2 analysis points at which efficacy will be considered:

- primary analysis of PFS, interim analysis of OS
- final analysis of OS

Approximately 524 patients will be enrolled in a 1:1 randomization (262 patients in the ramucirumab-plus-docetaxel arm and 262 patients in the placebo-plus-docetaxel arm). The primary analysis of PFS will be performed when a minimum of 331 PFS events have been observed from the first 437 randomized patients (that is, 24% censoring rate). If full enrollment is not reached at the point when 331 PFS events have been observed, the primary analysis will be performed when full enrollment is reached, on the number of PFS events observed upon full enrollment from the first 437 randomized patients. Although PFS is the primary endpoint, this study is adequately powered on OS. Assuming an OS hazard ratio of 0.75, this sample size yields at least 80% statistical power to detect superiority of the ramucirumab-plus-docetaxel arm over the placebo-plus-docetaxel arm using stratified log-rank test with a 2-sided type I error of 0.05.

A gatekeeping design will be used to assess PFS, OS, and ORR. The PFS, OS, and ORR are tested in a fixed sequential manner. The OS superiority is tested only if the PFS superiority test is significant. Likewise, the ORR superiority is tested only if the OS superiority test is significant. For OS, an interim analysis for efficacy is planned at time of PFS primary analysis; the Bonferroni method will be used for alpha splitting, with  $\alpha=0.001$  (2-sided) spent at the interim analysis and  $\alpha=0.049$  (2-sided) spent at the final analysis for OS.

The sample size to test PFS superiority at the PFS primary analysis is determined based on the following assumptions:

- the PFS hazard ratio for treatment group (ramucirumab plus docetaxel) vs. control group (placebo plus docetaxel) is 0.70.
- the randomization ratio is 1:1.
- a 2-sided significance level of 0.05.
- control arm median PFS = 2 months.
- the type II error rate is 0.1 (that is, the power of the trial is set to 90%).
- Enrollment rate is approximately 36 patients per month. This includes an estimated 25% chance of screen failure patients.

Under these assumptions, the estimated total number of PFS events at the PFS primary analysis is 331, from an expected accrual of 371 patients. Assuming a 15% patient dropout rate (to account for patients who will not be able to provide PFS events due to early dropout or meeting certain PFS censoring rules) from PFS follow-up, it is estimated that approximately 437 patients are needed to reach 331 PFS events in approximately 18.4 months. To ensure that the PFS data are sufficiently mature at the time of the primary PFS analysis, this analysis will be confined to the first 437 patients randomized, after these patients have generated 331 events.

The sample size to test OS superiority was determined based on the following assumptions:

- $\alpha = 0.049$  (2-sided), statistical power = 80%
- the randomization ratio is 1:1
- control arm median OS = 9 months
- hazard ratio = 0.75
- interim efficacy analysis for OS (at the PFS primary analysis) with testing at  $\alpha = 0.001$  (2-sided)

Under these assumptions, the estimated total number of OS events at the final OS analysis is 382 from an expected accrual of 497 patients. Assuming a 5% dropout rate (to account for patients who will not be able to provide OS events due to early dropout for reasons such as withdrawal of consent or lost to follow-up) from survival follow-up, it is estimated that 524 patients will need to be randomized to reach 382 OS events in approximately 33 months.

At the PFS primary analysis, the first 437 randomized patients will be used for the analysis. An interim OS analysis will be performed; all patients who have been randomized at that point will be used to compare OS between the 2 arms, based on Bonferroni alpha splitting of 0.001 (2-sided).

At the OS primary analysis, all randomized patients will be used to compare OS between the 2 arms. The type I error spend for OS is 0.049 (2-sided), which approximately corresponds to observed hazard ratio (HR)  $\leq 0.82$ .

The estimated number of events and efficacy boundaries and probabilities of first crossing the boundaries at each analysis are summarized in [Table JVDC.5.1](#). EAST<sup>®</sup> 6.3 is used to calculate the sample size and boundaries.

**Table JVDC.5.1. Study Design and Operating Characteristics**

Statistical Analysis		Primary PFS	Final OS
Duration of study (months)*		18.4	32.6
Estimated	# events	331	382
	# patients randomized	437	524
	P value boundary	$p \leq 0.05$	$p \leq 0.049$

Abbreviations: OS = overall survival; PFS = progression-free survival.

\* Estimated numbers.

## 5.2. General Considerations

This document describes the statistical analyses planned prior to final treatment assignment unblinding of the aggregate database. Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol and the justification for making the change will be described in the clinical study report (CSR). Additional exploratory analyses of the data will be conducted as deemed appropriate.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and all confidence intervals (CIs) will be given at a 2-sided 95% level, unless otherwise stated. Statistical analysis will be performed using SAS software (SAS, Version 9.1.2 or higher).

The following general terms will be used globally in the SAP:

- Unless otherwise specified, summary statistics stand for n, mean, standard deviation, median, Q1, Q3, minimum, and maximum for continuous variables; and frequency and percentage for categorical variables.
- **Study Treatment Period:** begins on the day the first dose of study treatment is administered and ends when the patient and the investigator agree that the patient will no longer continue study treatment. The date of this agreement is to be reported on the case report form (CRF) as the Date of Discontinuation from study treatment.
- **Postdiscontinuation Follow-Up:** begins the day after the patient and the investigator agree that the patient will no longer continue study treatment.

*Short-term follow-up* begins the day after the patient and the investigator agree that the patient will no longer continue study treatment and lasts approximately 30 days ( $\pm 7$  days) (until the short-term 30-day safety follow-up visit is completed).

*Long-term follow-up* begins the day after short-term follow-up is completed and continues until the patient's death or overall study completion (whichever is earlier).

### 5.2.1. Definitions of Analysis Variables

Definitions of efficacy, safety, and PRO analysis variables are listed in Section 5.2.1.1, Section 5.2.1.2, and Section 5.2.1.3, respectively. Other variables are listed below alphabetically.

- **Age (years):**  $(\text{Informed Consent Date} - \text{Date of Birth} + 1)/365.25$ .  
**Note:** Average days in a year = 365.25, reflecting the Julian Year of 3 years with 365 days each and 1 leap year of 366 days. Birth month and day are imputed to be July 1 because only birth year is collected through CRF.
- **Baseline measurement**
  - Efficacy: The last measurement on or prior to the date of randomization will serve as the baseline measurement. In the event such a value is missing, the first assessment completed prior to the first study drug administration will be used as the baseline assessment, so long as this assessment was taken within 7 days of randomization.
  - Safety: The last non-missing measurement prior to the first study drug administration will be used as the baseline assessment.
  - Demographic and other baseline characteristics: The last measurement on or prior to the date of randomization will serve as the baseline measurement. In the event such a value is missing, the first assessment completed on or prior to the date of first study drug administration will be used as the baseline assessment.
  - Patient-report outcome: The last measurement on or prior to the date of randomization will serve as the baseline measurement. In the event such as value is missing, any assessment prior to or on date of first study drug administration will serve as the baseline assessment.
- **Duration** is calculated as:
  - Duration (days):  $(\text{End Date} - \text{Start Date} + 1)$
  - Duration (weeks):  $(\text{End Date} - \text{Start Date} + 1)/7$
  - Duration (months):  $(\text{End Date} - \text{Start Date} + 1)/30.4375$   
**Note:** Days in months =  $(1/12) \times \text{average number of days in a year}$
  - Duration (years):  $(\text{End Date} - \text{Start Date} + 1)/365.25$
- **Duration of disease** is defined as months from first diagnosis of cancer to randomization.
- **Measurable disease (Yes/No)** is defined as yes for patients with at least 1 target lesion and no otherwise, based on radiographic assessment data collected at baseline.

- **Study Day:** Study day indicates the number of days the patient has been receiving study treatment. It is calculated as assessment date – first dose date + 1 day if the assessment is done on or after the first dose day. If the assessment is done prior to the first dose day, study day will be calculated as assessment date – first dose date. Date of first dose is defined as study day 1.

### 5.2.1.1. Efficacy Analysis Variables

Definition of efficacy analysis variables are listed.

**Progression-free survival** is defined as the time measured from the date of randomization to the date of radiographic documentation of progression (as defined by RECIST v.1.1) or the date of death due to any cause, whichever is earlier. The detailed censoring rule is provided in [Table JVDC.5.2](#).

**Table JVDC.5.2. Censoring Rule of PFS Primary Analysis**

Situation	Event / Censor	Date of Event or Censor
<b>Tumor progression or death</b>	Event	Earliest date of Progressive Disease (PD) or death
<b>No tumor progression and no death</b>	Censored	Date of last adequate radiological assessment or date of randomization (whichever is later)
<i>Unless</i>		
<b>No baseline</b> radiological tumor assessment available	Censored	Date of randomization
<b>No adequate post baseline</b> radiological tumor assessment available <u>and</u> death reported after 2 scan intervals following randomization	Censored	Date of randomization
<b>New anticancer treatment</b> started <u>and</u> no tumor progression or death within 14 days	Censored	Date of adequate radiological assessment prior to (start of new therapy +14 days) or date of randomization (whichever is later)
<b>Tumor progression</b> or death documented <u>immediately after</u> 2 or more consecutive missing scan intervals following last adequate radiological tumor assessment or randomization (whichever is later)	Censored	Date of last adequate radiological assessment prior to the missing assessment or date of randomization (whichever is later)

Abbreviations: CR = complete response; PD = progressive disease; PFS = progression-free survival; PR = partial response; SD = stable disease.

Notes:

Adequate radiological tumor assessment refers to an assessment with one of the following responses: CR, PR, SD or PD.

The 2-scan interval is counted from the date of last adequate tumor assessment to the date of next two scheduled tumor assessment plus 14 days (adjusted by tumor assessment window).

Refer to flow chart in [Appendix 1](#) if a patient meets multiple censoring criteria.

If there are multiple dates associated with one assessment, the assessment date will be set to the first date when the overall response is PD and the last date otherwise.

Table JVDC.5.3 lists censoring rules for sensitivity analysis (SA) definitions. Progression-free survival (day) = Date of progression / censor - Date of randomization + 1.

**Table JVDC.5.3. Censoring Rules for PFS Sensitivity Analysis Definitions**

Sensitivity Analysis Definition #	Situation	Date of Progression or Censor	Censored / Progressed
SA 1: count symptomatic deterioration as progression	Radiographic documented progression or symptomatic deterioration	Date of documented progression or date of symptomatic deterioration, whichever occurred first.	Progressed
SA 2: Ignore new anticancer treatment	New anticancer treatment (systemic therapy) started before radiographic documented progression or death	A) date of radiographic documentation of progression or death, whichever is earlier B) last adequate radiological assessment if no radiographic documented progress or death occurred	A) Progressed B) Censored
SA 3: Ignore missing tumor assessment	Death or radiographic documented progression after $\geq 2$ consecutively missed tumor assessment visits	Date of radiographic documentation of progression or death, whichever is earlier	Progressed
SA 4: Treat lost to follow up as progression	Patient is lost to follow-up without radiographic documented progression or death	Date of next scheduled post baseline radiological assessment at or after becoming lost to follow-up	Progressed

Abbreviations: PFS = progression-free survival; SA = sensitivity analysis.

**Overall survival (OS)** is defined as the time from the date of randomization to the date of death from any cause. If the patient was alive at the cutoff for analysis (or was lost to follow-up), OS data will be censored for analysis on the last date the patient was known to be alive.

**Objective response rate (ORR)** is defined as the proportion of randomized patients achieving a best overall response of complete response (CR) or partial response (PR) per RECIST v.1.1. Patients who do not have any post baseline tumor response assessments are considered non-responders and are included in the denominator when calculating the response rate.

**Note:** Tumor assessments performed after initiation of new anticancer treatment (systemic therapy) or after cancer-directed surgery following randomization will be excluded from evaluating the best overall response. Best overall response is the best response recorded from randomization until disease progression, in the order of CR, PR, stable disease (SD), and progressive disease (PD). Refer to Attachment 5 of the protocol for definitions of CR, PR, and SD.

**Disease control rate (DCR)** is defined as portion of randomized patients achieving a best overall response of CR, PR, or SD per RECIST v.1.1. Patients who do not have any post baseline tumor response assessments for any reason are considered non-responders and are

included in the denominator when calculating the response rate. Tumor assessments performed after initiation of new anticancer treatment (systemic therapy) or after cancer-directed surgery following randomization will be excluded from evaluating the best overall response.

**Duration of response (DOR)** is defined as the duration from the date of first evidence of a CR or PR to the date of radiographically documented progression or the date of death due to any cause, whichever is earlier. If a responder is not known to have died or have radiographically documented progression as of the data inclusion cutoff date, DOR will be censored at the date of the last adequate tumor assessment. This is defined for responders only.

**Time to progression (TTP)** is defined as the time from the date of randomization to the date of objectively determined PD. For patients who died without objective PD (including death from study disease), TTP was censored at the date of last adequate radiological assessment or date of randomization (whichever is later). For patients who were still alive at the time of analysis, and who did not have PD, TTP was censored at the date of the last objective progression-free disease assessment.

#### 5.2.1.2. Safety Analysis Variables

Definitions of variables for safety analysis are listed by category and alphabetically within category.

**Adverse event (AE)-related variables** are listed below:

- **Adverse event (AE)** is defined as any untoward medical occurrence in a patient administered a pharmaceutical product, without regard to the possibility of a causal relationship.
- **Adverse events of special interest (AESIs)** include arterial thromboembolic events (ATE), bleeding / hemorrhage (also gastrointestinal [GI] hemorrhage as a subcategory), congestive heart failure (CHF), fistula (GI and non-GI), GI perforation (non-Fistula), healing complication, hypertension, infusion-related reaction (IRR), liver injury/failure, proteinuria, reversible posterior leukoencephalopathy syndrome (RPLS), and venous thromboembolic events (VTE).

**Notes:**

Categories of AESI may be modified as the understanding of the safety of the investigational drug increases. The final list of Medical Dictionary of Regulatory Activities (MedDRA) terms will be maintained at both compound and study level and reported in the CSR.

- **Consolidated AEs** are composite AE terms consisting of synonymous preferred terms (PTs) to allow meaningful interpretation of the AE data. Consolidated AE categories and PTs will be maintained at compound and/or study level and reported in the CSR.
- **Serious adverse event (SAE)** is any AE that results in one of the following outcomes:

- death
- a life-threatening experience (that is, immediate risk of dying)
- persistent or significant disability/incapacity
- initial or prolonged inpatient hospitalization
- congenital anomaly/birth defect
- considered significant by the investigator for any other reason
- **Treatment-emergent adverse event (TEAE)** is defined as an AE that either emerges or worsens at any time between the first dose of study treatment and the end of the short term follow up visit after treatment discontinuation, regardless of causal relationship with the treatment.

**Exposure-related variables** are listed below:

- **Number of dose level reductions:** Sum of the number of dose level reductions as reported in the electronic CRF (eCRF)
- **Dose delays:** As reported in the eCRF
- **Dose omission:** As reported in the eCRF

**Ramucirumab or placebo treatment:**

- Duration of treatment (weeks; 21 days added to duration of treatment because administration is every 3 weeks [on Day 1 of each 3-week cycle]) = [(Date of last cycle day 1 – date of first dose) + 21] ÷ 7
- Cumulative dose, dose intensity, relative dose intensity:
  - Cumulative dose (mg/kg) = Sum of (dose administered at each infusion [mg] ÷ last available weight [kg])
  - Weekly dose intensity (mg/kg/week) = (Cumulative dose) ÷ (Duration of Treatment [week])
  - Planned weekly dose intensity (mg/kg/week) = 10 mg/kg / 3 weeks = 3.3 mg/kg/week
  - Relative dose intensity (%) = (Weekly dose intensity) ÷ (Planned weekly dose intensity) x 100%

**Docetaxel treatment:**

- Duration of treatment (weeks) = [(Date of last cycle day 1 – date of first dose) + 21] ÷ 7
- Cumulative dose, dose intensity, relative dose intensity:
  - Cumulative dose (mg/m<sup>2</sup>) = Sum of (dose administered at each infusion [mg] ÷ Last available body surface area (BSA) [m<sup>2</sup>])



- Weekly dose intensity (mg/kg/week) = (Cumulative dose) ÷ (Duration of treatment)
- Planned weekly dose intensity (mg/m<sup>2</sup>/week) = 75mg/m<sup>2</sup>/ 3 weeks = 25 mg/m<sup>2</sup>/week (60 mg/m<sup>2</sup>/3 weeks = 20 mg/m<sup>2</sup>/week for countries with 60 mg/m<sup>2</sup> starting dose)
- Relative dose intensity (%) = (Weekly dose intensity) ÷ (Planned weekly dose intensity) x 100%

### 5.2.1.3. Patient-Reported Outcome Analysis Variables

Definitions of variables for PRO analyses are listed below alphabetically.

#### EORTC QLQ-C30

The quality-of-life (QoL) assessment will be conducted using the EORTC-QLQ-C30, a self-administered, cancer-specific questionnaire with multidimensional scales.

The 30 items (Q1-Q30) of the QLQ-C30 are scored to obtain 15 scales (1 global health status/QoL scale, 5 functional scales, and 9 symptom scales/items). A linear transformation is used to derive scores ranging from 0 to 100 where:

- A high score for a functional scale represents a high / healthy level of functioning
- A high score for the global health status / QoL represents a high QoL
- A high score for a symptom scale / item represents a high level of symptomatology / problems

The following variables will be derived for each scale / single item:

- Change from baseline is calculated by subtracting baseline assessment result from the current assessment result.

Each scale/item at every assessment point will be compared to its baseline value and be categorized as follows:

- **Deterioration:** Defined as an increase of  $\geq 10$  points for the symptom scales or a decrease of  $\geq 10$  points for the functional scales and the global health status/QoL scale. For the time-to-deterioration analysis, a patient has sustained deterioration only if there is no subsequent on-treatment assessment without deterioration. That is, if any subsequent assessment returns to baseline, improves from baseline or is  $< 10$  points worse than baseline, then it is not counted as deteriorated. If the patient does not have any further assessments after the one associated with deterioration, then it is considered a deterioration event. If a scale deteriorates, then stabilizes or improves, and then deteriorates with no subsequent stabilization or improvement, then the second deterioration would count as a deterioration event. A deterioration event can be observed from any on-treatment assessment or at short-term follow-up. However, the short-term follow-up assessment can not be considered for any subsequent improvement or stabilization.

- Improvement: Defined as a decrease of  $\geq 10$  points for the symptom scales or an increase of  $\geq 10$  points for the functional scales and the global health status/QoL scale.
- Stable: Defined as no change or increase/decrease  $< 10$  points.

The ceiling effect is defined as the percentage of patients who have a baseline score of  $< 10$  points on the symptom scales, and  $\geq 90$  points on the functional scales and global health status/QoL scale. The floor effect is defined as the percentage of patients who have a baseline score of  $\geq 90$  points on the symptom scales, and  $< 10$  points on the functional scales and global health status/QoL scale.

### EQ-5D-5L

Each patient completing the EQ-5D-5L reports the level, or score for each of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), as well as a visual analog scale (VAS) score. The index score is calculated as a function of individual levels (1=no problem, 2=slight, 3=moderate, 4=severe, and 5=extreme problem) responses from each of the 5 dimensions based on country-specific algorithms. The index score will not be computed for an assessment if the patient has 1 or more missing values among the 5 items. The index score is based on a scale of 0 (death) to 1 (perfect health), but scores  $< 0$  are possible based on the algorithm. The value set for England will be used, with range of scores of -0.281 to 1 (Devlin et al. 2016 [<https://www.ohe.org/publications/valuing-health-related-quality-life-eq-5d-5l-value-set-england>]). The VAS is a score reported by the patient ranging from 0 (worst imaginable health state) to 100 (best imaginable health state).

The following variables will be derived for the EQ-5D-5L index and VAS:

- For each patient, change from baseline will be calculated for every postbaseline assessment by subtracting the baseline assessment result from the current assessment result.

### **5.3. Adjustments for Covariates**

As supportive analysis, the primary and secondary efficacy endpoints will also be analyzed adjusting for pre-specified potential prognostic factors chosen from the variables listed below. Detailed description as for which factors to be used will be provided for relevant analyses in later sections.

- Randomization stratification factors per eCRF:
  - Eastern Cooperative Oncology Group (ECOG) performance status (0 vs. 1)
  - geographic region (North America, Europe/Other, or East Asia), where the “Europe/Other” region includes the remainder of countries to which patients are enrolled [list countries], apart from those in the “North America” region (includes United States, Canada, Mexico) and the “East Asia” region (includes Japan, Republic of Korea [also known as South Korea], and Taiwan)
  - visceral metastasis (yes vs. no), where visceral metastases involve the liver, lung, and/or bone

- Other factors of interest:
  - sex (males vs. females)
  - age subgroup (<65 vs. ≥65)
  - race (White vs. Asian vs. all others)
  - prior adjuvant therapy (neo-adjuvant vs. adjuvant vs. no prior adjuvant therapy)
  - prior platinum-based therapy (carboplatin vs. cisplatin vs. other)
  - prior non-platinum-based therapy (gemcitabine vs. methotrexate/vinblastine/doxorubicin vs. other)
  - prior immune checkpoint inhibitor (Yes vs. No)
  - number of metastatic sites (≤2 vs. ≥3)
  - liver metastasis (Yes vs. No)
  - disease progression within 14 months after the last dose of adjuvant or neoadjuvant therapy. Patients will be grouped into 2 subgroups: relapsed <6 months vs. relapsed ≥6 months.
  - number of regimens administered for the treatment of metastatic disease (0 vs. ≥1)
  - weight (>75 kg vs. ≤75 kg)
  - hemoglobin (<10 g/dL [ $<100$  g/L] vs. ≥10 g/dL [ $\geq 100$  g/L])
  - primary tumor location (bladder vs kidney/ureter/urethra)
  - pathology (mixed pathology [squamous cell, small cell, other] vs. pure transitional cell)

#### 5.4. Handling of Dropouts or Missing Data

Rules for handling dropouts or missing data are listed by type of analysis alphabetically. Unless otherwise specified, observed data will be used and missing data will not be imputed or carried forward.

##### General rules for imputing dates related to AE or concomitant therapy:

- Onset date of an AE or start date of a concomitant therapy:
  - If only the day is missing, the date will be set to:
    - First day of the month that the event occurred, if the onset yyyy-mm is after the yyyy-mm of first study treatment.
    - The day of the first study treatment, if the onset yyyy-mm is the same as yyyy-mm of the first study treatment.
  - If both the day and month are missing, the complete date will be set to:

- January 1 of the year of onset, if the onset year is after the year of the first study treatment.
- The date of the first dose, if the onset year is the same as the year of the first study treatment.
- Resolution date of an AE or end date of a concomitant therapy
  - If only the day is missing, the date will be set to the last day of the month of the occurrence.
  - If both the day and month are missing, the date will be set to December 31 of the year of occurrence. If a date is completely missing, then the AE will be considered treatment-emergent. In case of additional therapies, the therapy will be considered concomitant.

**General rule for imputing other dates:** If a date variable is needed for an analysis, use the following general rule to impute incomplete date:

- If the date has no missing year and month but the day is missing, then assign Day 1 to the day
- If the date has no missing year, but has missing month, then assign January to the month.

However, after imputation, check if the imputed date is logically consistent with other relevant date variable(s) and make appropriate correction if necessary. For example, if a visit start date was May 10, 2008 and a tumor assessment date was May xx, 2008 (missing day) but it was known that it occurred after that visit, then after imputation, the tumor assessment date became May 01, 2008. In this case, the imputed tumor assessment date should be compared to the visit start date and then corrected to be the visit start date, May 10, 2008.

**Patient-reported outcome analysis:** For percentage compliance of the EORTC QLQ-C30 and EQ-5D-5L, instruments with at least 1 item completed will be considered as having been completed. When scoring scales of the QLQ-C30, missing items can be imputed if  $\geq 50\%$  of the remaining items in the scale are non-missing (Fayers et al. 2001). No other adjustment or imputation for missing data will be performed.

**Safety analysis:** The following rule for missing data processing will apply for safety analysis:

- Missing classifications concerning study medication relationship will be considered as related to study medication (both components).
- If the AE onset date is missing or partial, the date will be compared as far as possible with the date of first dose of study medication when determining whether or not the AE is present at baseline. In this case, the AE will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the AE started prior to the first dose of study medication.

**Time-to-event analysis:** All censored data will be accounted for using appropriate statistical methods. See Section 5.2.1 and Section 5.11 for details.

## 5.5. Multicenter Studies

This is a multicenter, randomized, double-blind study. Investigative center was not a stratification factor because the large number of investigative centers would breakdown the intended balance within each combined stratification level by the stratified randomization method. It will not be included as a covariate in any covariate-adjusted analysis because the large number of investigative centers in this study cannot be practically incorporated into such analysis.

In anticipating a regulatory submission in Japan, subgroup analyses of Japan are planned. There is no other plan for any subgroup analyses within a particular center or group of centers. However, retrospective exploratory analyses of center-specific data or region based subgroup analysis may be conducted as deemed appropriate.

## 5.6. Multiple Comparisons/Multiplicity

The gatekeeping group sequential design based on alpha spending functions controls the type-1 error for the key efficacy endpoints including PFS, OS, and ORR. Other safety-only interim analyses are also planned which will not affect the type-1 error for the primary analysis.

## 5.7. Study Patients

The following summaries (frequency and percentage) and listings for patient disposition will be performed:

- Patient disposition by investigator site and country and overall: patients entered (that is, signed informed consent), entered but not randomized, randomized (that is, intent-to-treat [ITT] population), randomized but not treated, treated (that is, safety population), in Per-Protocol population, and eligible for PRO analysis by (refer to Section 5.7.1 for analysis population definitions)
- Reasons for discontinuation for the following patients groups:
  - all treated patients
  - screen fail patients (that is, patients who entered but not randomized)
  - randomized patients who did not receive any study treatment
- Listings of
  - primary reason for discontinuation
  - date of randomization, first dose administration, last dose administration, and treatment discontinuation

### 5.7.1. Analysis Populations

Table JVDC.5.4 lists analysis population definitions and associated data type for analysis.

**Table JVDC.5.4. Analysis Populations**

<b>Population</b>	<b>Definition</b>	<b>Analysis Type / Variable</b>	<b>Note</b>
<b>Intent-to-treat PFS population</b>	The first 437 randomized patients in this trial	PFS (inv and ind), PFS sensitivity analyses (per <a href="#">Table JVDC.5.3</a> ), ORR/DCR (inv and ind), DOR (inv and ind), TTP (inv and ind), Demographics, baseline disease characteristics	Patient will be grouped according to randomized treatment.
<b>Intent-to-treat (ITT) Population</b>	All randomized patients	Demographics, baseline characteristics, OS, PROs	Patients will be grouped according to randomized treatment.
<b>Safety Population (SP)</b>	All randomized patients who received any quantity of study drug	Safety, e.g. dosing/exposure, AE, hospitalizations, concomitant medication and transfusions	Patients will be grouped according to treatment received as defined by the first dose received.
<b>Per-Protocol Population (PPP)</b>	ITT patients who received any study therapy and did not have any important protocol deviations that could potentially affect the efficacy conclusion. These important protocol deviations are detailed in <a href="#">Section 5.7.2</a> .	PFS first 437 randomized and OS (as a supportive analysis); other efficacy endpoints may also be analyzed as needed	PPP will only be performed if there are significant numbers of patients with important protocol deviations ( $\geq 5\%$ of total ITT population).

Abbreviations: AE = adverse event; DCR = disease control rate; DOR = duration of response; ind = independently assessed; inv = investigator assessed; ITT = intent-to-treat; PFS = progression-free survival; PRO = patient-reported outcome; ORR = objective response rate; OS = overall survival; TTP = time to progression.

Compliance for the PRO instruments will be reported for the ITT population. Time to deterioration and proportions of patients with improved, worsened, or stable scores will be reported for the ITT population. All other PRO analyses will be based on the ITT population and will include those from whom a completed PRO instrument was obtained at baseline and at least one post-baseline (either during study treatment period or short-term follow-up period); thus the actual patients included for each analysis will depend on the instrument and PRO variable.

A patient listing of analysis population details will be provided. This listing will be presented by treatment group and will include: investigator site, patient identifier, inclusion/exclusion flag for each population and reason for exclusion from each population. All patients screened will appear on this listing.

### **5.7.2. Important Protocol Deviations**

The Per Protocol population is a subset of the ITT population and consists of the randomized and treated patients who do not have an important protocol deviation (that is, clinically important and potentially impact efficacy evaluations) including but not limited to as listed below:

- patients who were unblinded prior to primary analysis of PFS
- patients who do not have histopathologically confirmed, locally advanced or unresectable or metastatic urothelial (transitional cell) carcinoma of bladder urethra, ureter, or renal pelvis (or as allowed in inclusion criterion 1)
- patients who did not receive prior platinum-based therapy
- patients who have not had a prior immune checkpoint inhibitor and demonstrated disease progression greater than 14 months after completing first-line platinum regimen
- patients that have had a prior immune checkpoint inhibitor and demonstrated disease progression greater than 24 months after completing first-line platinum regimen
- patients who have received more than one prior systemic chemotherapy regimen for metastatic disease (except as allowed in Inclusion Criteria). A treatment regimen must consist of minimum of 2 cycles to be considered as a prior regimen.
- patients with baseline ECOG performance status (ECOG PS) score  $\geq 2$
- patients with untreated brain metastases at baseline
- patients who have received additional concurrent anticancer therapy, biological response modifiers, other investigational agents and/or radiation therapy (except for palliation to symptomatic sites of disease) while receiving study treatment.
- patients who received incorrect study dose at least 25% of the times (that is, number of incorrect infusions/total number of infusions patient received  $\geq 25\%$ ). For an individual infusion dose to be considered incorrect, it would have to deviate  $>10\%$  from the correct dose.

The list of patients (except for patients with incorrect study medication) included in this population will be identified prior to unblinding for the primary PFS analysis.

The protocol deviations described in this SAP are restricted to deviations that can be assessed by analysis of data available in the clinical database.

### 5.8. Demographic and Other Baseline Characteristics

The following patient demographic and other baseline characteristics will be summarized:

- patient demographics: age (years) and age groups ( $<65$  vs.  $\geq 65$ ; and  $65-<75$ ,  $75-<85$ ,  $\geq 85$  years), gender, ECOG PS, race, ethnicity, height (cm), weight (kg), BSA ( $m^2$ )
- potential prognostic factors as listed in Section 5.3
- baseline disease characteristics:
  - at initial diagnosis only: disease stage
  - at study entry only: duration of disease (months)

- disease location: bladder, renal, ureter, urethra
- prior cancer therapies: type of therapy (surgery, radiotherapy, systemic therapy), type of prior surgery, type of prior radiotherapy, type of prior systemic therapy
- prior platinum-based therapy (carboplatin vs. cisplatin vs. other)
- Historical illness (no versus at least one diagnosis) by MedDRA PT, presented in decreasing frequency  
**Note:** Subjects reporting more than one condition/diagnosis within a PT will be counted only once for that PT
- Comparison between the eCRF and IWRS values of the stratification factors

Patient listings of demographic data and baseline characteristics will be provided. Patient listings of prior cancer therapies (surgery, radiotherapy, and systemic therapy) will be provided.

## 5.9. Concomitant Medications

The following concomitant medications used in the study treatment period or the 30-day post discontinuation follow-up period will be summarized by numbers and percentages by treatment group, presented in decreasing frequency of the World Health Organization (WHO) drug term across treatment arms:

- all concomitant medications
- premedication for study drug

Select classes of drugs (for example, granulocyte colony-stimulating factors, antiemetics) based on consolidated WHO drug terms will be summarized by treatment arm.

The proportions of patients reporting use of concomitant medications will be compared between the treatment groups. Patient listing of all concomitant therapies and premedications will be provided.

## 5.10. Treatment Compliance

Ramucirumab, placebo, and docetaxel will be intravenously administered only at the investigational sites. As a result, patient compliance is ensured.

## 5.11. Efficacy Analyses

### 5.11.1. Primary Efficacy Analyses

The primary analysis of PFS will be performed when a minimum of 331 PFS events have been observed from the first 437 randomized patients (that is, 24% censoring rate). If full enrollment is not reached at the point when 331 PFS events have been observed, the primary analysis will be performed when full enrollment is reached, on the number of PFS events observed upon full enrollment from the first 437 randomized patients.



For the primary comparison of PFS, the primary endpoint, between the assigned study treatment arms, a stratified log-rank test at alpha level of 0.05 (two-sided) will be performed to test the following statistical hypotheses about the PFS HR for ramucirumab over placebo:

$H_0$ : PFS HR  $\geq 1.00$  (ramucirumab not superior to placebo)

$H_a$ : PFS HR  $< 1.00$  (ramucirumab superior to placebo).

The stratification will be the same as that used for randomization. An unstratified log-rank test will also be performed.

The following analyses of PFS will also be performed:

- Summary of PFS events (number and percentage), censoring rate, and reasons for censoring
- Restricted Mean Difference Analysis

The common method for describing benefit on the time scale is to calculate the difference in median event time between the 2 treatment arms. An alternative method for describing benefit on the time scale is to estimate the signed difference between the Kaplan-Meier (KM) curves. This corresponds to calculating the difference in the average time to event for the 2 treatment arms (Irwin 1949; Karrison 1997; Meier et al. 2004). Similar to the HR, this method uses all of the available information across the KM curves, but has the additional advantage of assessing benefit on the time scale.

To estimate an improvement in PFS with ramucirumab, we will follow the method of Irwin (1949) detailed in Karrison (1997) and Meier et al. (2004) for estimating the “difference in average PFS,” which we will refer to more formally as the restricted mean difference in PFS. The area under each KM curve will be calculated using numerical integration (trapezium rule) per Karrison and implemented in SAS using PROC LIFETEST. The difference between treatment arms and a CI for the difference will be formed.

Since the KM curve may be ill-determined beyond a certain range, or even undefined (if the longest observation is censored), for evaluation and comparison of means, the area under each KM curve will be calculated between time 0 and restriction time T, which is why this is referred to as a “restricted mean.” Following the suggestion of Karrison, the restriction time T will be chosen as largest time point t such that the standard error (SE) of the survival estimate at time t in each treatment group is no more than 0.075. For this purpose, we will use the simple, albeit conservative, formula proposed by Peto et al. (1977) for calculating the SE of S(t) as  $SE(S(t)) = S(t)\sqrt{((1-S(t))/n(t))}$ , where n(t) is the number of patients still at risk at time t.

- Kaplan-Meier survival curve (Kaplan and Meier 1958) by treatment group will be provided.

- The Kaplan-Meier method will be used to estimate parameters (medians, quartiles, and percentages), difference of percentages and associated 95% CIs and p-values for landmark analyses on each treatment group at 3, 6, 9, 12, and 24 months. Patients who did not have the event at the corresponding time point will be considered right-censored observations.
- Hazard ratio for treatment effect will be estimated using Cox proportional hazards (PH) model stratified identically to the primary log-rank test with assigned treatment as the only covariate, reported with 2-sided 95% CIs and Wald's test p-value. This Cox PH model will be referred to as the primary Cox PH model henceforth.

### 5.11.2. Secondary Efficacy Analyses

#### 5.11.2.1. Supportive Analyses of Primary Efficacy Endpoint

The following supportive analyses of PFS will be performed:

- HR for treatment effect will be estimated using an unstratified Cox PH model
- as a sensitivity analysis, the primary PFS analysis will be repeated using stratification based on the eCRF values.

Hazard ratio for treatment effect will be estimated using a multivariate Cox PH model to be constructed by selecting covariates among all the variables listed in Section 5.3 using stepwise selection method. Treatment will not be included with the selected covariates. The stepwise selection will use an entry p-value  $<0.05$  and exit p-value  $\geq 0.10$ . Once the final model with the selected covariates has been determined, the treatment factor will be added and the treatment effect (hazard ratio) estimated along with the corresponding 95% CI .

**Note:** If a level of a covariate consists of fewer than 5% of randomized patients, analysis within that level may be omitted.

- As a sensitivity analysis, the primary PFS analysis will be repeated for the per-protocol population.
- As sensitivity analyses, the primary PFS analysis will be repeated using different PFS censoring rule as defined in Table JVDC.5.3, to evaluate whether and to what extent the conclusion of the PFS analysis under the primary definition would be affected under the different censoring rules.

#### 5.11.2.2. Key Secondary Efficacy Analyses

A gatekeeping design will be used to assess PFS, OS, and ORR. The PFS, OS, and ORR are tested in a fixed sequential manner. The OS superiority is tested only if the PFS superiority test is significant. Likewise, the ORR superiority is tested only if the OS superiority test is significant. For OS, an interim analysis for efficacy is planned at time of PFS primary analysis; the Bonferroni method will be used for alpha splitting, with  $\alpha=0.001$  (2-sided) spent at the interim analysis (as long as the primary PFS test is significant), and  $\alpha=0.049$  (2-sided) spent at the final analysis for OS.

For OS, the same analyses used for the analyses of the primary endpoint PFS (except sensitivity analyses based on PFS censoring rule) will be performed.

If OS analysis is significant, then testing on ORR will be conducted at same alpha level. Objective tumor response (CR+PR) rate (ORR) will be reported along with exact confidence bounds (CI: 95%) and compared using the Cochran-Mantel-Haenszel test adjusting for the stratification variables. A sensitivity analysis on ORR will be performed to consider only confirmed response. A sensitivity analysis on ORR based on independent assessment may be conducted.

Investigator assessments of radiological data will be considered primary for the analysis of PFS. An independent central review of radiologic data (lesion measurements based on computed tomography [CT] scans or magnetic resonance imaging [MRI]) will be conducted. Analyses will be performed examining the agreement/discordance between the investigator assessment and the independent assessment data. Progression-free survival analysis of the independent assessment data may be performed and serve as a sensitivity analysis. This analysis will be based on the same patients as the analysis for investigator-reviewed data (that is, the first 437 randomized patients at the primary PFS analysis).

#### **5.11.2.3. Analyses of Other Efficacy Endpoints**

Disease control (CR+PR+SD) rate (investigator assessment and independent assessment) will be reported along with exact confidence bounds (CI: 95%) and compared using the Cochran-Mantel-Haenszel test adjusting for the stratification variables.

Time to progression (investigator assessment and independent assessment) will be compared between both treatment groups using stratified log-rank test and Kaplan-Meier estimates.

Duration of response (investigator assessment and independent assessment) will be compared between both treatment groups using unstratified log-rank test and Kaplan-Meier estimates. This analysis is for responders only.

A waterfall plot of the (best) percent change from baseline in target lesion measurement will be generated for each study arm.

Patient listings of tumor assessments (target and non-target lesion assessments and tumor response), PFS investigator and independent assessments, and OS will be provided.

#### **5.11.3. Additional Overall Survival Analyses**

During the PFS final/OS interim analysis, a delayed OS treatment effect was observed with the OS K-M curves not separating until after 6 months. Exploratory analyses indicated that this phenomenon may be attributable to potentially differing treatment effects according to some predefined subgroup analyses such as baseline hemoglobin, no prior therapy in the locally advanced/metastatic setting, as well as the Bellmunt risk score (3-factor score incorporating ECOG performance status, liver metastases, and hgb), a composite risk index known to be prognostic in urothelial carcinoma (Bellmunt et al. 2010). The Bellmunt risk score showed that those patients with a total score of 2 or 3 had an OS of 3.8 and 1.7 months, respectively. One of

the inclusion factors of the RANGE study required a life expectancy of at least 3 months. It is possible that those patients enrolled with a total Bellmunt score of 2 or 3 account for many of the early events leading to delayed OS K-M separation. Of greater importance, patients with a life expectancy in the 1.7-to-3.8-month range may not benefit from additional therapy. Patients with lower Bellmunt risk scores, patients with baseline Hgb $\geq$ 10 g/dL and have had at least one prior therapy in the locally advanced/metastatic setting, and primary tumor site in the bladder may have the potential to achieve greater treatment effects with ramucirumab.

Given these observations, an additional analysis will be added to the end of the testing order. Per the current SAP, final OS will be tested in the ITT population with an alpha level of 0.049. The testing order will preserve the order of testing in the current SAP-specified gatekeeping approach (PFS, OS, and ORR). Then the subgroup, patients with Bellmunt risk score  $<2$  (from the 3-factor risk score), will be tested for OS at an alpha level of 0.049, using only the independent increment of the OS information (that is, OS events observed after the data cut off for the primary PFS analysis) to control for type 1 error (Jennison and Turnbull 1999).

This additional test on a subgroup will be performed subsequent to the test of ORR, so as to preserve the order of testing PFS, OS, and ORR in the gatekeeping approach as currently specified in the SAP. Note that using only the independent increment (that is, new events observed after the primary PFS analysis) of the OS information in this subgroup analysis properly controls the familywise type 1 error. Technical details of the independent increment of survival times can be found in [Appendix 2](#).

Exploratory analyses for final OS:

The independent increment (new survival events observed after the primary PFS analysis) will be tested for the 2 subgroups listed below, sequentially, in the order given:

1. For the subgroup of patients with at least 1 prior regimen in the locally advanced/metastatic setting and Hgb  $\geq$ 10 g/dL
2. For the subgroup of patients with primary tumor site in the bladder.

Additional subgroup analyses of OS in the ITT subpopulation will be conducted on patients with:

1. Bellmunt risk scores  $< 2$ , and
2. At least 1 prior regimen in the locally advanced/metastatic setting and Hgb  $\geq$ 10 g/dL,
3. Primary tumor site of bladder.

Given the delayed OS treatment effect observed as OS interim analysis, for the final OS test on the ITT population, a  $\lambda$ -family weighted log rank test ( $\lambda = 0$ ,  $\lambda = 1$ ), which weighs late events more than early events, will be added as a supportive analysis (Fleming and Harrington 1991). This weighted log rank test will have more power if there is a delayed treatment effect.

#### **5.11.4. Subgroup Analyses**

Progression-free survival and OS HRs for treatment effect and corresponding 95% CIs will be estimated using the unstratified Cox PH model for each of the subgroups listed in Section 5.3. A forest plot of the estimated HRs and their 95% CIs will be provided. If the number of events in a particular subgroup is less than 15, this subgroup may not be presented in forest plot.

Additional subgroup analyses may be performed as deemed appropriate. The goal of subgroup analyses is to assess internal consistency of study results, and whether there is significant treatment heterogeneity across any of the subgroups. Appropriate interpretation is important since, even if all patient subgroups benefit to exactly the same extent in truth, smaller or larger estimated effects, even negative effects, may be seen for some subgroups simply by chance alone. Without appropriate interpretation, this can lead to erroneous conclusion in one or more subgroups, in particular where differential treatment effects are not expected across any of the factors assessed. In order to assist with interpretation of the subgroup results, the methodology of Fleming (1995) will be followed to provide background information on the extent of variability that might be expected by chance alone if deemed necessary.

The limitations of subgroup become more severe as the subgroup size gets smaller, and at a certain point, results from small subgroups may provide no value but mislead.

#### **5.12. Post-Discontinuation Therapy**

The numbers and percent of patients reporting post-discontinuation therapies (PDTs) will be provided overall, and by type of therapy (surgery, radiotherapy, or systemic therapy). Surgery and radiotherapy will be further characterized by intent. Systemic therapy will be further categorized by WHO drug terms.

Additional analysis may be explored for helping interpret OS results, for example time to PDT.

#### **5.13. Patient-reported Outcome Analyses**

For each instrument, percentage compliance will be calculated as the number of completed assessments divided by the number of expected assessments (that is, patients still on study). Percentage compliance will be summarized by treatment group and overall. Reasons for noncompliance will be summarized.

##### **5.13.1. EORTC QLQ-C30**

###### Descriptive Summaries

Data will be summarized descriptively for all of the 15 scales at each assessment time by treatment arm, provided at least 30% of the patients have assessments at that assessment time.

Floor and ceiling effects will be summarized for each of the 15 scales.

###### Response Analysis:

The proportion of patients in each arm with deterioration in scores, improvement in scores, stable scores, and missing scores (due to progression, death, or failure to complete) at specific time

points will be presented for each of the 15 scales in a summary table. The denominator should include all patients in the ITT population. The proportion of patients who improved or have stable scores at each assessment time point will be compared using Fisher's exact test.

#### Time to Deterioration:

For each of the 15 scales, the time to sustained deterioration, calculated as randomization date to date of first sustained deterioration (as described in Section 5.2.1.3), will be compared between the treatment groups using the log-rank test. The HR and 95% CIs will be estimated using Cox PH model with assigned treatment and baseline score as covariates.

### **5.13.2. EQ-5D-5L**

Summary statistics for the 5 dimensions, index, and VAS will be calculated for each assessment period by treatment group.

## **5.14. Safety Evaluation**

### **5.14.1. Exposure**

The following exposure-related variables will be reported using summary statistics (number of patients, mean, and standard deviation) by treatment group:

- exposure: duration of treatment; number of cycles received; number of patients completing  $\geq 1$  cycle,  $\geq 2$  cycles, ...,  $\geq 6$  cycles, and mean, SD; number of patients with dose adjustments: dose reduction, dose delay, and dose omission;
- reasons for dose adjustments

The following exposure-related variables will be reported using summary statistics (number of patients, mean, SD, median, 1st and 3rd quartiles, minimum, and maximum) by treatment group:

- dose intensity: cumulative dose; weekly dose intensity; relative dose intensity, overall weekly dose intensity, overall relative dose intensity

Details of study drug administration will be included in patient listings.

### **5.14.2. Adverse Events**

The most current version of MedDRA at time of analysis will be used when reporting AEs by MedDRA terms. Unless otherwise specified, when summarized by PT, AEs will be presented in decreasing frequency of PT based on Arm A; when summarized by System Organ Class (SOC) and PT, AEs will be presented in decreasing frequency of PT within SOC based on Arm A. If more than 1 AE is recorded for a patient within any SOC or PT term, the patient will only be counted once on the most severe grade and the closest relationship to treatment.

#### **5.14.2.1. Overall Summary of Adverse Events**

An overall summary of AEs will be provided to summarize the following categories using frequency counts and percentages:

- patients with at least 1 TEAE, treatment-emergent serious adverse event (TE-SAE), Common Terminology Criteria for Adverse Events (CTCAE) Grade 3 and above
- patients with AEs that led to death (on study, within 30 days after treatment discontinuation),
- patients with AEs that led to discontinuation and SAEs that led to discontinuation

The summary will be provided for regardless of study drug causality, and repeated for events deemed by the investigator to be related to study treatment. Comparison between the treatment groups will be performed using Fisher's exact test.

#### **5.14.2.2. Treatment-Emergent Adverse Events**

The following summaries of TEAEs will be provided (\*repeat for events deemed by the investigator to be possibly related to study medication, †include consolidated summary):

- by PT\*†
- >5% in the experimental arm by PT\*†
- by SOC and PT\*
- by maximum CTCAE grade and by PT\*†

A patient listing of all AEs will be provided.

#### **5.14.3. Deaths, Serious Adverse Events, and Other Significant Adverse Events**

Reasons for deaths (study disease, AE [any study drug-related, investigational drug-related, procedural related]) will be summarized separately for 1) all deaths, 2) on study treatment deaths, within 30 days after treatment discontinuation, and more than 30 days after treatment discontinuation.

Serious adverse events will be summarized by SOC and PT, by PT and repeated for events deemed by the investigator to be possibly related to study medication, with consolidated summary performed if needed. A listing of SAEs will be produced.

In addition, the following analyses will be performed (\*repeated for events deemed by the investigator to be possibly related to study medication, †include consolidated summary):

- adverse events leading to death by PT†
- adverse events leading to study treatment discontinuations by PT†
- adverse events leading to study treatment dose modification by PT†
- Adverse events of special interest by PT\*
- liver injury/failure
- listing of deaths
- listing of AESIs

#### **5.14.4. Clinical Laboratory Evaluation**

The severity of laboratory results will be classified according to CTCAE Version 4.0. The shifts in CTCAE toxicity grading from baseline to worst grade postbaseline (first dose up to 30 days after the last dose of study treatment) will be produced.

A patient listing of all laboratory data will be provided with a flag for values outside of the laboratory normal range as well as investigator site, patient identifier, age, gender, race, and visit.

#### **5.14.5. Hospitalizations and Transfusions**

The frequency and percentage of patients with any hospitalizations experienced during the study treatment period or 30-day post discontinuation follow-up period will be summarized by treatment group. Hospitalization incidence rates will be compared between the treatment groups using Fisher's exact test. In addition, total number of days in hospital and admissions will be summarized and compared using the Wilcoxon rank sum test. These will be further characterized by reason (study-drug-related, non-study-drug-related).

The frequency and percentage of patients with any blood transfusions experienced during the study treatment period or 30-day post discontinuation follow-up period will be summarized by treatment group. Transfusions will be further characterized by transfused blood product (for example, packed red blood cells, platelets, fresh frozen plasma, or whole blood).

#### **5.14.6. Vital Signs, Physical Findings, and Other Observations Related to Safety**

A summary of ECOG PS at each scheduled time point will be provided. Actual value and change from baseline for vital sign measurements will be summarized at each assessment time point using summary statistics. Electrocardiogram (ECG) measurements will be summarized at each assessment time point using summary statistics. Listings of ECOG PS, vital signs, ECG data will be provided.

#### **5.14.7. Subgroup Analyses**

Analyses will be performed for the following safety subgroups:

- age (<65 years versus  $\geq 65$  years)
- race (White versus Asian versus Other)
- gender (females versus males)
- geographic region (North America versus Europe/Other versus EA),

Analyses for these subgroups will include:

- summary of patient disposition
- dose intensity of study drugs
- reasons for dose adjustments and dose delays



- overview of AEs
- treatment-emergent adverse events summarized by PT and consolidated terms
- Common Terminology Criteria for Adverse Events v4 Grade 3 and above
- serious adverse events summary by PT and consolidated terms
- adverse events of special interest by PT and consolidated terms
- reasons for deaths
- adverse events leading to study treatment discontinuation by PT
- adverse events leading to study treatment dose modification summarized by PT
- listing of SAEs

## 5.15. Pharmacokinetics and Immunogenicity

### 5.15.1. Pharmacokinetics

#### Ramucirumab

Serum concentrations of ramucirumab prior to infusion (minimum observed drug concentration [ $C_{\min}$ ]) and post-end of ramucirumab infusion (approximately maximum observed drug concentration [ $C_{\max}$ ]) will be summarized using descriptive statistics. Additional analysis utilizing the population PK approach may also be conducted if deemed appropriate.

Relationships between ramucirumab exposure and measures of efficacy and safety will be explored. Detailed exposure-response analysis plan is described separately.

#### Docetaxel

Plasma concentrations of docetaxel at each time point will be summarized using descriptive statistics and presented by docetaxel dose level.

### 5.15.2. Immunogenicity

For immunogenicity, the number and percent of patients with positive ramucirumab antibody response will be summarized. Additional efficacy or safety analyses may be performed in the subgroup of patients with positive ramucirumab antibody response. The antibody response and any alteration in ramucirumab PK may also be explored, as well as any relationship with experiencing an infusion reaction. Further exploratory analyses may be performed as appropriate.

## 5.16. Translational Research

Translational research analyses will be performed according to a separate analysis plan.

### 5.17. Interim Safety Analysis

An independent data monitoring committee (IDMC) will be established prior to first patient visit (FPV). The IDMC will review unblinded interim analyses of safety. Study sites will receive information about interim results ONLY if they need to know for the safety of their patients.

All interim analyses will include complete assessments of safety-related data. Patient enrollment will continue during the conduct of these analyses. The interim analyses will be performed as follows:

- The first safety data review will occur after at least 100 evaluable patients in each arm have started treatment in Cycle 3 or discontinued all study treatment due to any reasons prior to Cycle 3.
- The second IDMC meeting will occur after tallying approximately 250 evaluable patients who have either started treatment in Cycle 3 or discontinued all study treatment prior to Cycle 3 for any reason.
- Additional IDMC meetings may occur if the IDMC requests, or if safety concerns arise during routine safety monitoring for this study.

The details on the process flow/communication plan among study team, IDMC and senior management are provided in the IDMC charter.

The following analyses will be performed for safety interim analysis, with analysis population as specified in [Table JVDC.5.4](#).

#### Demographic and Other Baseline Characteristics

- patient demographics

#### Patient Disposition

- patient disposition
- reasons for treatment discontinuation as well as patients continuing on the study

#### Exposure

- summary statistics for exposure-related variables
- dose intensity of study drugs
- reasons for dose adjustments and dose delays

#### Adverse Events

- overview of AEs
- treatment-emergent adverse events TEAEs summarized by PT\*
- Common Terminology Criteria for Adverse Events v 4.0 Grade 3 and above
- serious adverse events summarized by PT\*
- adverse events of special interest by PT\*
- reasons for deaths

- adverse events leading to study treatment discontinuations summarized by PT
- adverse events leading to study treatment dose modification summarized by PT
- listing of SAEs
- listing of preexisting conditions and AEs

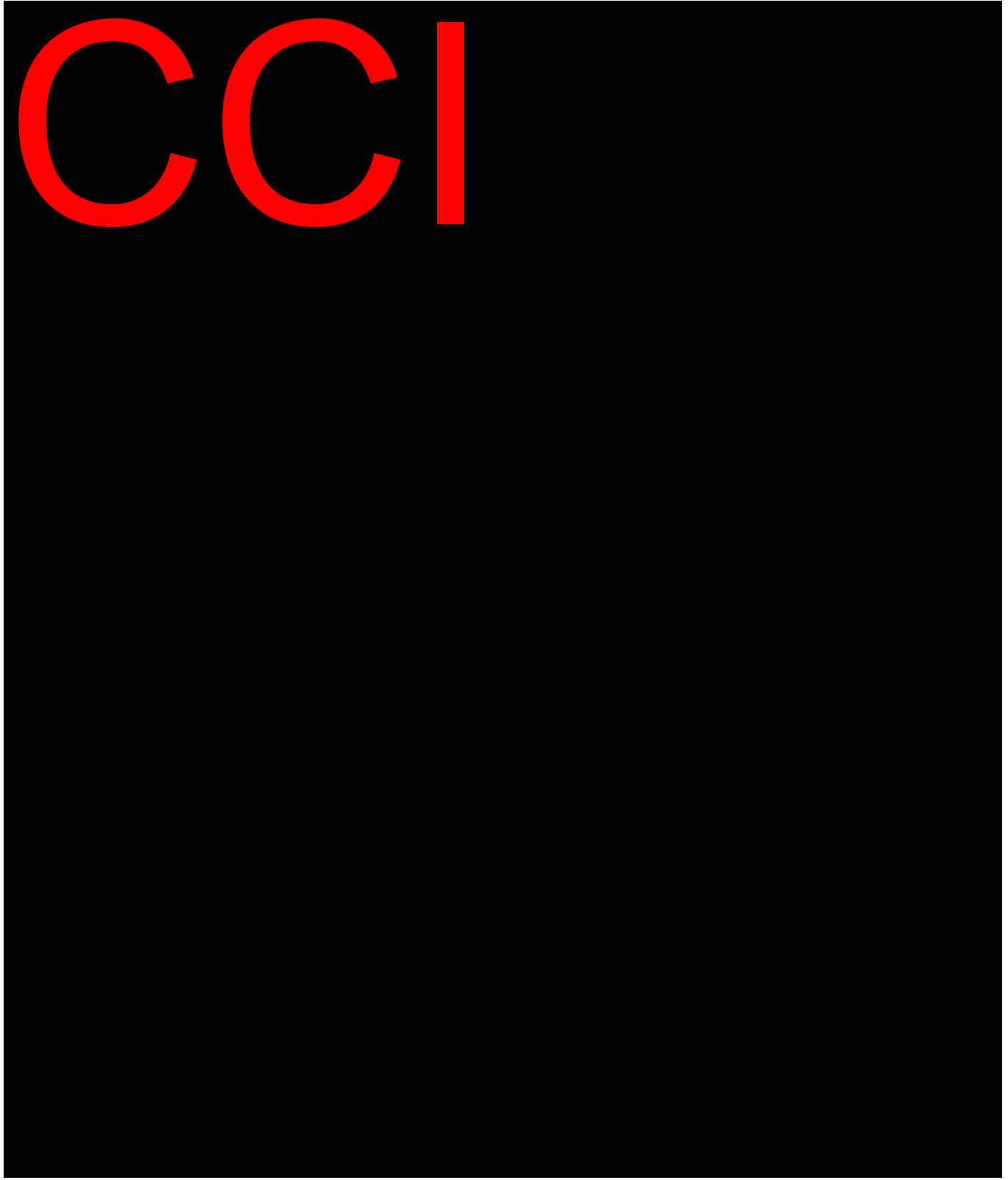
\*Repeat for events deemed by the investigator to be possibly related to study medication.

### **5.18. Clinical Trial Registry Analyses**

For the purpose of fulfilling the Clinical Trial Registry (CTR) requirements, summary of SAEs (whether treatment emergent or not) and ‘Other’ AEs (that is, non-serious TEAEs) by PT and treatment group will be performed. For each PT, the number of patients at risk, patients who experienced the event, and events will be presented. In addition, the summary will be provided as a dataset in XML format.

## 6. Unblinding Plan

This unblinding plan refers to the process to be followed for the primary PFS analyses.

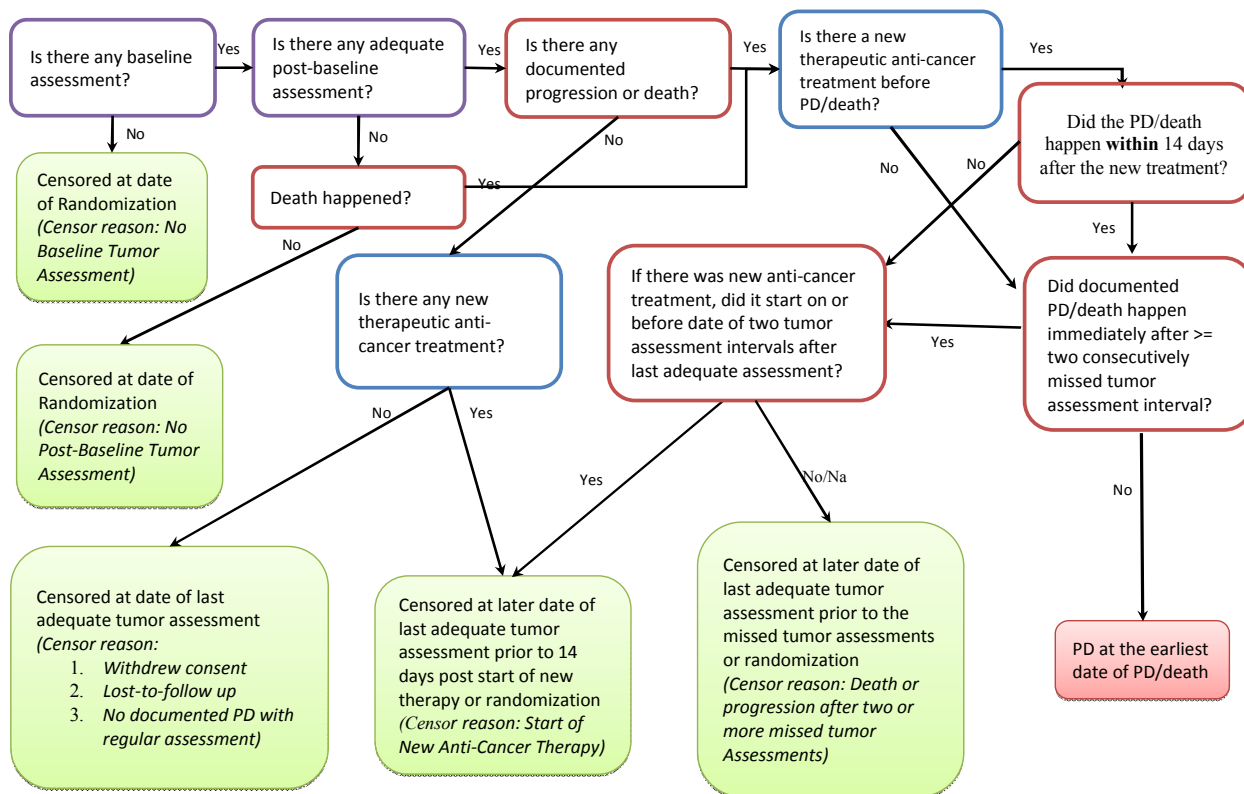




## 7. References

- Bellmunt J, Choueiri TK, Fougerey R, Schutz FAB, Salhi Y, Winquist E, Culine S, von der Maase H, Vaughn DJ, Rosenberg JE. Prognostic factors in patients with advanced transitional cell carcinoma of the urothelial tract experiencing treatment failure with platinum-containing regimens. *J Clin Oncol*. 2010;28:1850-1855.
- Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. 2016. Available at: <https://www.ohe.org/publications/valuing-health-related-quality-life-eq-5d-5l-value-set-england>. Accessed January 10, 2017.
- Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC; 2001.
- Fleming TR. Interpretation of subgroup analyses in clinical trials. *Drug Inf J*. 1995;29:1681S-1687S.
- Fleming TR, Harrington DP. Counting processes and survival analysis. New York: John Wiley; 1991.
- Irwin JO. The standard error of an estimate of expectational life. *J Hygiene*. 1949;47:188-189.
- Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials. New York: Chapman & Hall; 1999.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
- Karrison TG. Use of Irwin's restricted mean as an index for comparing survival in different groups--interpretation and power considerations. *Control Clin Trials*. 1997;18(2):151-167.
- Meier P, Karrison T, Chappell R, Xie H. The price of Kaplan-Meier. *J Am Stat Assoc*. 2004;99(467):890-896.
- Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br J Cancer*. 1977;35(1):1-39.

## Appendix 1. Flow Chart of PFS Censoring Rules



---

## Appendix 2. Technical Details on Independent Increment of Survival Times

---

Lilly proposes to modify the Statistical Analysis Plan (SAP) to incorporate additional analyses on OS.

Final OS will still be tested on the ITT population with an alpha level of 0.049 per the current SAP. If the test is significant, OS will be tested on the subgroup of patients with Bellmunt scores < 2 at an alpha level of 0.05, using **only the independent increment** of the OS information to control for type 1 error (Jennison and Turnbull 1999). More technical details are provided below:

Let  $d_T$  and  $d_C$  be the number of events for treatment and control arm.

Let  $\tau_1 \cdots \tau_{d_T+d_C}$  be the events times. Let  $\delta_i$  be the event indicator at time  $\tau_i$ .

Let  $r_{i,T}$  and  $r_{i,C}$  be the at-risk set at time  $\tau_i$ .

Then  $\delta_i$  follows a hypergeometric distribution with mean

$$e_i = \frac{r_{i,T}}{r_{i,T} + r_{i,C}}$$

and variance

$$v_i = \frac{r_{i,T}r_{i,C}}{(r_{i,T} + r_{i,C})^2}$$

Thus, the standard log rank statistics compares each  $\delta_i$  and  $e_i$  at each event time in the treatment arm, and is defined as

$$Z = \frac{\sum_{i=1}^{d_T+d_C} (\delta_i - e_i)}{\sqrt{\sum_{i=1}^{d_T+d_C} v_i}}.$$

Now denote  $I = \sum_{i=1}^{d_T+d_C} v_i$ .

We can then calculate  $z^{(1)}$  and  $z^{(2)}$  for the OS analysis at OS interim (PFS final) and OS final with their responding  $I^{(1)}$  and  $I^{(2)}$ . But define

$$\tilde{z}^{(2)} = \frac{\sqrt{I^{(2)}}z^{(2)} - \sqrt{I^{(1)}}z^{(1)}}{\sqrt{I^{(2)} - I^{(1)}}}.$$

Then  $\tilde{z}^{(2)}$  is independent of  $z^{(1)}$ , which will be our OS test statistics of the gated-subgroup.



Leo Document ID = f4f15cef-124b-4f84-9a3a-017a6319746c

Approver: PPD

Approval Date & Time: 13-Feb-2018 17:34:43 GMT

Signature meaning: Approved