Janssen Research & Development

Statistical Analysis Plan
Amendment 2

A Phase 3 Study of Erdafitinib Compared With Vinflunine or Docetaxel or Pembrolizumab in Subjects with Advanced Urothelial Cancer and Selected FGFR Gene Aberrations

Protocol 42756493BLC3001; Phase 3; Amendment 6

JNJ-42756493 (Erdafitinib)

Status:	Approved
Date:	24 January 2023
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-ERI-164733590, 3.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

VERSION HISTORY

Statistical Analysis Plan (SAP) Version History Summary

SAP Version	Issue Date
Original SAP	24Sep2018
Amendment 1	31Mar2021
Amendment 2	24Jan2023

Amendments below are listed beginning with the most recent amendment

Amendment 2 (24 Jan 2023)
Summary of Key Changes: Per FDA's request via email dated September 8, 2021, added references to the final SAP to justify the assumptions for the OS medians used in sample size calculation.

Sections	Description of Changes	Rationale
Section 1.2.1. Figure 1. Schematic Overview of the Study	Added a note clarifying Long-term Extension (LTE) Phase is not included in the schematic but available in Protocol Amendment 6, Attachment 8.	Per changes in Protocol Amendment 6 that LTE phase was added.
Section 3. Sample Size Determination and Appendix 8	Added references to justify the assumptions for OS medians used in sample size calculations for each cohort.	To address feedback from the Food and Drug Administration (FDA).
Section 5.6.4, Table 5 Categorical variables used for subgroup analysis.	Added the term "territories" to align with Asian Pacific (APAC) marketing reference guide.	Alignment with APAC marketing reference guide.
Section 5.5.2.1. Treatment-Emergent Adverse Events	Removed summary tables of TEAEs with incidence higher than 10% and the corresponding 5% summary table is kept. Revised the paragraph for TEAEs subgroup analysis.	TEAEs with incidence frequency higher than 10% are covered in the summary table of TEAEs with incidence higher than 5%. TEAEs will be summarized for selected subgroups as necessary.
Section 5.5.3.5. Amsler Grid Test	Revised this section to include Ophthalmologic Examinations	To be consistent across the erdafitinib program.
Section 5.5.2.2. Anticipated/Unlisted (unexpected) Adverse Event	Removed the section.	Anticipated events will be handled by a separate unblinded team.
Section 6.6. Appendix 6 Adverse Events of Special Interest and Clinical Interest	Revised to add new preferred terms for adverse events of special interest and clinical interest.	To be consistent with the most current version of the Medical Dictionary for Regulatory Activities Terminology (MedDRA) dictionary.
5.1.1. Level of Significance	Designated TUSD-3 as PRO secondary endpoint.	To address feedback from the Food and Drug Administration (FDA) Type C meeting that TUSD-3 composite score is most relevant for this study. (FDA correspondence reference ID 4260896)
5.4.1.3. PRO-Related Endpoints	Added TUSD-3 and other PRO endpoints will be analyzed and added "psychometric	To clarify additional details of analyses of TUSD-3 and other PRO

Amendment 2 (24 Jan 2023)
Summary of Key Changes: Per FDA's request via email dated September 8, 2021, added references to the final SAP to justify the assumptions for the OS medians used in sample size calculation.

Sections	Description of Changes	Rationale
	analysis plan (PAP)" as a separate analysis plan for PRO-related endpoints.	endpoints are also available in a separate psychometric analysis plan.
5.6.2.1. Patient-reported Outcomes (PROs) Assessments	Dropped the term "Assessment" from the title "Patient-reported Outcomes (PROs) Assessments".	For clarification.
5.6.2.1.1. FACT-BL	Changed FACT-BL 36 items to 39 items.	FACT-BL contains 39 items, not 36 items.
5.6.2.2. Analysis Methods	Deleted the sentence "Depending on the emerging data from the qualitative study (Protocol Number: JN1027C v0.1), a separate PRO SAP might be generated."	PRO PAP was developed.
5.6.2.2.2. Change Scores from Baseline Over Time	Deleted the term "utility" from EQ-5D-5L utility scores. Replaced "swim plot" with means (±SD) plots and LS means (±SE) plots	Change scores will also be calculated for the EQ-5D VAS, not just utility values Swim plot does not apply to change from baseline analysis
5.6.2.2.3. Missing Data Handling	5.6.2.2.3. Missing Data Handling	No missing data handling will be implemented as these are not standard practices.
5.6.2.2.4. Derivation of Meaningful Change thresholds	Removed details of the derivation of meaningful change thresholds	Already described in the PRO PAP

Amendment 1 (31Mar2021)

Summary of Key Changes: Per protocol amendment 5, the statistical analysis of the study has been revised to reflect the changes to the interim analysis (IA) timing for both cohorts.

Sections	Description of Changes
Section 1.1.1. Objectives and Endpoints Section 1.1.2.3. Exploratory Endpoints	Added an exploratory objective/endpoint to evaluate the relationship between CYP2C9 polymorphism and PK of erdafitinib. Removed the evaluation of the impact of study treatment on physical function (activity), mobility, and sleep because actigraphy is not performed in this study.
Section 3. Sample Size Determination	Changed the information fraction for IA to 65% for Cohort 1 and Cohort 2.
Section 5.1.1. Level of Significance	Added details of significant levels for primary and secondary endpoints at IA and final analysis for both cohorts utilizing a hierarchical testing approach: For both Cohort 1 and Cohort 2, the significance levels to be used at the IA and final analysis are 0.0108 and 0.0466, respectively. Replaced "TUSD-3" with "PROs (FACT-B1, Time Until Symptom Deterioration, PGIS and EQ-5D-5L)" in the secondary endpoints testing order.

Amendment 1 (31Mar2021)

Summary of Key Changes: Per protocol amendment 5, the statistical analysis of the study has been revised to reflect the changes to the interim analysis (IA) timing for both cohorts.

Sections	Description of Changes
5.3.2. Estimand	Added the section based on the SAP template, version 15.0 (see the note at the end of this table), to describe the estimand framework for the study including primary objective, study intervention, population, variable of interest and summary measure as well as intercurrent events and corresponding strategies.
Section 5.3.3. Analysis Methods Sections 5.3.3.1 and 5.3.3.2	Revised to indicate that the sample size re-estimation at IA no longer-applies to either cohorts and consequently removed sections 5.3.3.1 and 5.3.3.2 from the original SAP for the calculation method of the weighted Z statistics at final analysis.
Section 5.3.3.1.2	Added "Treatment crossover after disease progression may occur and its impact on OS may be explored."
Section 5.6.1	Added "Measurements of erdafitinib free concentration and free fraction may also be listed and summarized, if available."
5.6.3. Concordance of Central Laboratory and Local Historical test results of Selected FGFR Aberrations	Added per Independent Data Monitoring Committee's (IDMC) request to assess the concordance of the central laboratory and the local historical test results of selected FGFR aberrations.
Section 5.6.4, Table 5 Categorical variables used for subgroup analysis.	Added liver and lung metastases. Removed bladder cancer subtype.
Section 5.7. Interim Analysis	Updated to reflect that superiority and futility will be assessed at 65% information fraction for Cohort 1 and Cohort 2 at IA. The sample size adjustment guideline was removed.
Section 6.2, Table 7 Baseline Characteristics	Removed bladder cancer subtype.

Note: This SAP is reformatted according to the new template (Version 15.0 - 30 September 2020). Some reorganization and addition of some sections were implemented, such as the added Estimand section (Section 5.3.2) and the re-ordered sections including Subgroup Definition (Section 5.6.4) and Interim Analysis (Section 5.7).

1. INTRODUCTION

The phase 3 study aims to demonstrate superiority of single agent erdafitinib over established chemotherapy (docetaxel or vinflunine) or an anti-PD-(L)1 agent (pembrolizumab) in relapsed/refractory subjects with advanced urothelial cancer harboring selected FGFR aberrations who have progressed following at least 1 line of prior systemic therapy.

This document presents the Statistical Analysis Plan (SAP) for protocol 42756493BLC3001 (Amendment 5). This SAP contains the details and methods to be used to perform the proposed analysis of the safety, efficacy, pharmacokinetic (PK), and other secondary and exploratory endpoints.

This SAP is based upon the following study documents:

- Study Protocol, Amendment 6
- The latest version of the electronic Case Report Form (eCRF) (V15.0, 18Jan2022)

The SAP is in compliance with guidelines provided in the International Conference on Harmonization (ICH) Topic E9 Statistical Principles for Clinical Trials. In the event of future amendments to the protocol, this SAP may be modified as necessary to account for changes relevant to the statistical analysis.

1.1. Objectives and Endpoints

1.1.1. Objectives

Primary Objective

The primary objective of this study is to evaluate efficacy of erdafitinib versus chemotherapy or pembrolizumab in subjects with advanced urothelial cancer harboring selected FGFR aberrations who have progressed after 1 or 2 prior treatments, at least 1 of which includes an anti-PD-(L)1 agent (cohort 1) or 1 prior treatment not containing an anti-PD-(L)1 agent (cohort 2).

The primary endpoint of overall survival will be evaluated in 2 cohorts:

- Cohort 1: erdafitinib versus chemotherapy (docetaxel or vinflunine) [subjects who have received prior anti-PD-(L)1 agent]
- Cohort 2: erdafitinib versus pembrolizumab [subjects who have not received prior anti-PD-(L)1 agent]

The statistical analysis for the primary endpoint of overall survival and secondary efficacy endpoints will be performed independently for each cohort.

Secondary Objectives

- To evaluate progression-free survival (PFS) of subjects treated with erdafitinib versus chemotherapy or pembrolizumab.
- To evaluate the objective response rate (ORR) of subjects treated with erdafitinib versus chemotherapy or pembrolizumab.
- To evaluate the health-related quality of life (HRQOL) of subjects treated with erdafitinib chemotherapy or pembrolizumab.
- To evaluate the duration of response (DOR) for subjects treated with erdafitinib versus chemotherapy or pembrolizumab.
- To characterize the safety profile of subjects treated with erdafitinib versus chemotherapy or pembrolizumab.
- To evaluate the population pharmacokinetics (PK) of erdafitinib.

Exploratory Objectives

- To evaluate DNA, RNA, or protein biomarkers in tissue and blood samples which potentially predict tumor response or resistance to erdafitinib, chemotherapy, or pembrolizumab.
- To assess the expression of immune markers (eg, PD-(L)1) and determine molecular subtype in tumor samples.
- To evaluate changes in peripheral blood immune cell types, levels, and activation status in response to erdafitinib, chemotherapy, or pembrolizumab.
- To assess changes in tumor immune cell infiltrate and gene expression related to bladder cancer subtype, in response to erdafitinib in paired tumor biopsies.
- To evaluate the relationship between erdafitinib exposure and efficacy and safety endpoints.
- To evaluate the relationship between CYP2C9 polymorphism and pharmacokinetics (PK) of erdafitinib

1.1.2. Endpoints

1.1.2.1. Primary Endpoint

The primary endpoint is overall survival (OS), which is measured from the date of randomization to the date of the subject's death. If the subject is alive or the vital status is unknown, the subject will be censored at the date the subject was last known to be alive.

1.1.2.2. Secondary Endpoints

- PFS: duration in days from the date of randomization to the date of disease progression (or relapse from complete response [CR]) assessed per Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 by the investigator or death, whichever is reported first. For subjects who do not have disease progression and are alive, as well as for subjects with unknown disease progression or unknown survival status as of the clinical cutoff date, PFS will be censored at the date of the last adequate disease assessment. If there is no post-baseline tumor assessment for a subject, PFS will be censored on the date of randomization. Adequate disease

assessment is defined as having sufficient evidence to indicate correctly that progression has or has not occurred.

- ORR: the proportion of subjects who achieve complete response or partial response (PR), as assessed per RECIST v1.1 by the investigator.

- Change from baseline in patient-reported health status and physical functioning scales of the Functional Assessment of Cancer Therapy – Bladder Cancer (FACT-Bl), Time Until Symptom Deterioration (subset of FACT-Bl items), Patient-Global Impression of Severity (PGIS), and utility and visual analog scale of the European Quality of Life-5 Dimensions-5 Levels Questionnaire (EQ-5D-5L). (See a separate PRO SAP for details).

- DOR: for responders, duration in days from the date of initial documentation of a response to the date of first documented evidence of progressive disease (or relapse for subjects who experience CR during the study) or death. The censoring rule is similar to PFS.

- Safety: collection of adverse events (AEs), clinical laboratory values, electrocardiograms, vital signs, ophthalmologic evaluations, physical examinations.

- Oral clearance, area under the plasma concentration-time curve (and other parameters, as needed and as data permits) of erdafitinib will be estimated using a population PK approach.

1.1.2.3. Exploratory Endpoints

- PD-L1 expression level by immunohistochemistry (IHC), and bladder cancer subtype by RNA sequencing or other method(s) as appropriate

- Profile and activation status of peripheral immune cell subtypes by flow cytometry, T-cell receptor (TCR) sequencing, or alternate method(s) as appropriate

- Tumor T-cell infiltrate in paired biopsies by immunohistochemistry, and gene expression related to bladder cancer subtype by RNA sequencing or alternate method(s) as appropriate

- Use of models, such as Emax, to evaluate parameters describing the effect of erdafitinib exposure on clinical endpoints; for instance, erdafitinib concentration or metrics of exposure leading to 50% of the maximal effect

- Effect of CYP2C9 polymorphism on the PK of erdafitinib

1.2. Study Design

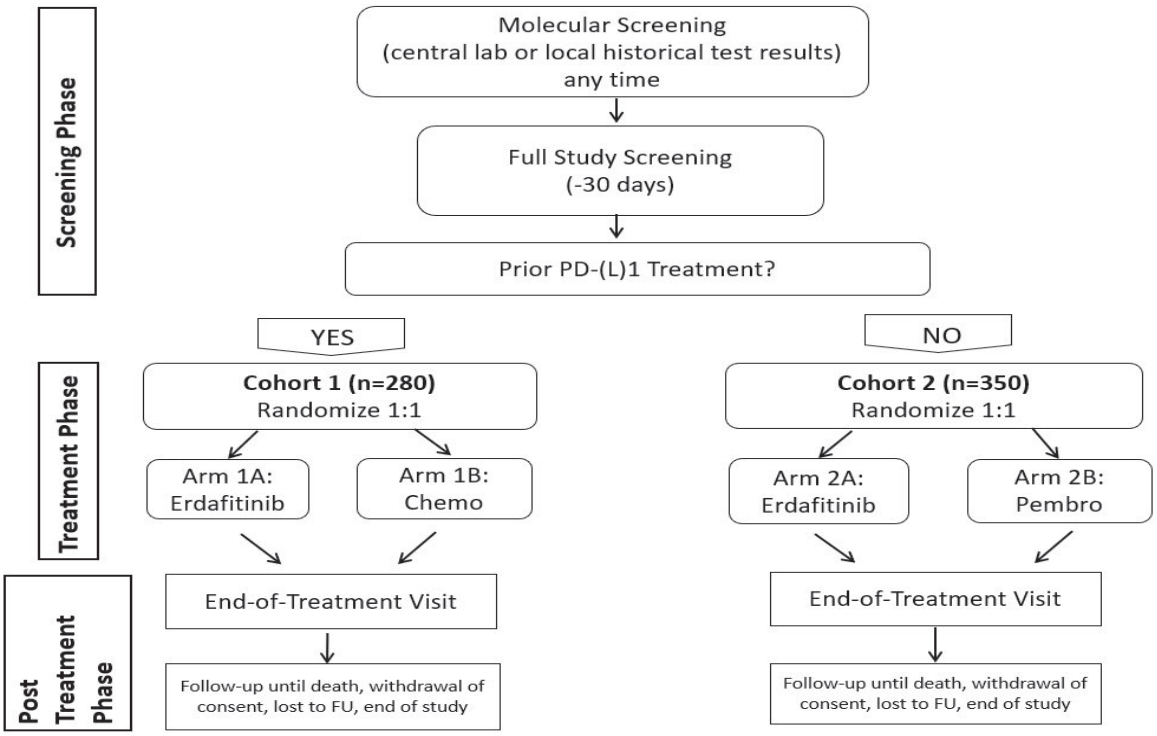
1.2.1. Introduction

This is a randomized, open-label, multicenter, global phase 3 study of erdafitinib versus standard of care, consisting of chemotherapy (docetaxel or vinflunine) or anti-PD-(L)1 agent pembrolizumab, in subjects with advanced urothelial cancer and selected FGFR aberrations who have progressed on or after 1 or 2 prior treatments (Cohort 1) or 1 prior treatment (Cohort 2).

Subjects will be assigned to Cohort 1 or Cohort 2 based upon prior treatment with an anti-PD-(L)1 agent. In Cohort 1, subjects who have received prior anti-PD-(L)1 will be randomized to erdafitinib versus chemotherapy (approximately 280 subjects). In Cohort 2, subjects who have not received prior anti-PD-(L)1 will be randomized to erdafitinib versus pembrolizumab (approximately 350 subjects). Cohort 1 and Cohort 2 will be assessed independently.

A diagram of the study design is provided below in Figure 1.

Figure 1: Schematic Overview of the Study



Note: The Long-term Extension Phase is not included in the schematic (see Protocol Amendment 6, Attachment 8).
*Treatment until disease progression, intolerable toxicity, withdrawal of consent, or decision by investigator.

1.2.2. Randomization and Blinding

Central randomization will be implemented in this study. Subjects will be assigned to Cohort 1 or Cohort 2 based upon prior treatment with PD-(L)1 agent. Within each cohort, subjects will be randomly assigned to 1 of 2 treatment groups based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor. In Cohort 1, subjects must have progressed on or after 1 or 2 prior line(s) of systemic treatment, with at least 1 line including anti-PD-(L)1 as monotherapy or as combination therapy, will be randomized 1:1 to erdafitinib (Arm 1A) or chemotherapy (vinflunine or docetaxel, Arm 1B). In Cohort 2, subjects must have progressed after one prior line of systemic treatment not containing anti-PD-(L)1 and will be randomized 1:1 to erdafitinib (Arm 2A) or pembrolizumab (Arm 2B). Stratified randomization will be implemented within each cohort with the following factors: region (North

America vs. European Union [EU] vs. Rest-of-the-World [ROW]), Eastern Cooperative Oncology Group (ECOG) performance status (0 or 1 vs. 2), and disease distribution (presence vs. absence of visceral [lung, liver, or bone] metastases). The interactive web response system (IWRS) will assign a unique treatment code, which will dictate the treatment assignment and matching study drug kit for the subject. The requestor must use his or her own user identification and personal identification number when contacting the IWRS and will then give the relevant subject details to uniquely identify the subject.

2. STATISTICAL HYPOTHESES

Cohort 1 hypothesis: Erdafitinib treatment prolongs OS in subjects with advanced urothelial cancer harboring selected FGFR aberrations following 1 or 2 prior line(s) of systemic therapy, with at least 1 line containing anti-PD-(L)1, compared with the OS of those treated with chemotherapy (docetaxel or vinflunine).

Cohort 2 hypothesis: Erdafitinib treatment prolongs OS in subjects with advanced urothelial cancer harboring selected FGFR aberrations following 1 prior line of systemic chemotherapy without anti-PD-(L)1, compared with the OS in those treated with pembrolizumab.

3. SAMPLE SIZE DETERMINATION

Cohort 1 will enroll approximately 280 subjects (approximately 140 subjects to each arm). The data cut-off date for the final analysis will be when approximately 208 death events have occurred. Assuming 53% improvement in median OS for the erdafitinib arm over the chemotherapy arm (a hazard ratio [HR] of 0.65 [Appendix 8] for the erdafitinib relative to chemotherapy group, under the exponential distribution assumption), the study has at least 85% power to detect a HR of 0.65 at a statistical significance level of 5% (2-sided) with one interim analysis for efficacy at an approximately 65% information fraction (approximately 136 deaths) and a final analysis.

Cohort 2 will enroll approximately 350 subjects (approximately 175 subjects to each arm). The data cut-off date for the final analysis will be when approximately 264 death events have occurred. Assuming 46% improvement in median OS for the erdafitinib arm over the pembrolizumab arm (a hazard ratio of 0.69 [Appendix 8] for the erdafitinib relative to the pembrolizumab group, under the exponential distribution assumption), the study has at least 85% power to detect a HR of 0.69 at a statistical significance level of 5% (2-sided) with one interim analysis for efficacy at an approximately 65% information fraction (approximately 172 deaths) and a final analysis.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Analysis Sets	Description
Intent-to-Treat (ITT)	The Intent-to-Treat (ITT) Analysis Set includes all randomized subjects. Subjects in this population will be analyzed according to the treatment to which they are randomized. The ITT Analysis Set will be used to summarize the study population and baseline characteristics, efficacy, PRO data, and health economic data.
Per Protocol (PP)	Per-Protocol (PP) Analysis Set includes all randomize subjects and meet all eligibility (inclusion/exclusion) criteria. If the PP analysis set comprises > 95% of the ITT analysis set, no analysis by PP will be performed.
Safety	The Safety Analysis Set includes all randomized subjects who received at least 1 dose of study drug. Safety data will be analyzed according to the actual treatment received. The Safety Analysis Set will be used to summarize the safety data, unless otherwise specified.
Pharmacokinetics-evaluable	The Pharmacokinetic-evaluable Analysis Set includes all randomized subjects who received at least 1 dose of erdafitinib and had at least 1 evaluable pharmacokinetic sample obtained posttreatment.
Biomarker	The Biomarker Analysis Set includes those randomized subjects with at least 1 biomaterial available and who have consented to participate in the study's biomarker evaluations.

5. STATISTICAL ANALYSES

5.1. General Considerations

Descriptive statistics and subject listings will be used to summarize the data. For continuous variables, number of observations, means, standard deviations, medians, quartiles, and ranges will be used. For discrete variables, frequency will be summarized. For time-to-event variables, Kaplan-Meier (KM) estimates will be provided. Additionally, a stratified log-rank test will be used to compare the survival curves within each cohort at a 2-sided alpha level of 0.05. The three stratification factors to be used in the analysis are as follows: region (North America vs. EU vs. ROW), ECOG performance status (0 or 1 vs. 2), and disease distribution (presence vs. absence of visceral [lung, liver, or bone] metastases). If necessary, data will be pooled for these stratification factors based on the rules defined in Section 5.1.2.

5.1.1. Level of Significance

The type 1 error will be controlled at 5% (2-sided) within each cohort. All tests will be conducted at a 2-sided alpha level of 0.05 and 95% confidence intervals (CIs) will be provided, unless stated otherwise.

For both Cohort 1 and Cohort 2, the significance levels to be used at the IA and the final analysis are approximately 0.0108 and 0.0466, respectively. The exact significance level at the interim analysis is to be determined by the observed number of events using the O'Brien-Fleming alpha-spending function implemented by the Lan-DeMets method.

A secondary null hypothesis will be tested within each cohort if and only if the primary null hypothesis and all the secondary null hypotheses that precede it have been rejected, i.e., utilizing a hierarchical testing approach as proposed by Tang and Geller (1999).

At both interim and final analyses, the secondary endpoints will be tested at the same significance levels as specified for testing the primary endpoint OS to protect the overall type I error rate. The testing order of these endpoints is as follows:

- PFS
- ORR
- Time to Urinary bladder cancer Symptom Deterioration (TUSD-3 as defined in Section 5.6.2.2.5)

If the null hypothesis for any of these endpoint fails to be rejected at the IA, then any subsequent endpoint(s) listed above will not be tested until the final analysis. If the null hypothesis for an endpoint is rejected at the IA, it will remain rejected and will not be re-tested at the final analysis.

5.1.2. Pooling Algorithm

The stratification factors to be used in the primary and major secondary efficacy analysis will be entered into the statistical models in the following order: ECOG performance status (0 or 1 vs. 2), disease distribution (presence vs. absence of visceral [lung, liver, or bone] metastases) and region (North America vs. EU vs. ROW), which separates the total subjects into 24 subgroups when treatment arm is taken into consideration. For time-to-event analyses, if any one of these subgroups has less than 10 corresponding events for any treatment arm, then the subjects will be pooled together by dropping stratification factors until there are at least 10 events in each stratum. The stratification factors will be dropped in the following order: region, disease distribution, ECOG.

5.2. Participant Dispositions

All enrolled subjects that entered (1) molecular screening and (2) full study screening will be summarized overall based on the following:

- Subjects with archived samples received
- Subjects who passed molecular screening
- Subjects who had full study screening
- Subject who passed full study screening

Subjects who failed screening or dropped out of the study before randomization will be summarized by the reasons.

The number of randomized subjects in the disposition categories will be summarized throughout the study by treatment group. A listing of subjects will be provided for the following categories:

- Subjects who discontinued study agent
- Subjects who terminated study prematurely

- Subjects who were randomized but did not receive study agent.
- Subjects who completed the study

5.3. Primary Endpoint Analysis

5.3.1. Definition of Endpoint

The primary efficacy endpoint, OS, is defined as the time from randomization to the date of death due to any causes. The OS (in months) is calculated as:

(date of death − date of randomization + 1)/30.4375.

If the subject is alive or the vital status is unknown (for example, lost to follow-up or withdrew consent etc.), the OS will be censored at the date the subject was last known to be alive. Subjects lacking data beyond randomization will have their OS censored at the date of randomization. ITT analysis set will be used for OS analyses.

5.3.2. Estimand

Primary Trial Objective: To evaluate efficacy of erdafitinib versus chemotherapy (vinflunine or docetaxel) or pembrolizumab in subjects with advanced urothelial cancer harboring selected FGFR aberrations who have progressed on or after 1 or 2 prior treatments, at least 1 of which includes an anti-PD-(L)1 agent (cohort 1), or 1 prior treatment not containing an anti-PD-(L)1 agent (cohort 2). The primary objective will be independently assessed within each cohort.

Study intervention:

- Erdafitinib
- Chemotherapy (Cohort 1) or pembrolizumab (Cohort 2)

Population: Subjects with advanced urothelial cancer harboring selected FGFR aberrations who have progressed on or after 1 or 2 prior treatments, at least 1 of which includes an anti-PD-(L)1 agent (Cohort 1), or 1 prior treatment not containing an anti-PD-(L)1 agent (Cohort 2).

Variable: Overall survival (OS)

Summary measure (Population-level summary): HR of erdafitinib vs. chemotherapy or pembrolizumab

Intercurrent events and their corresponding strategies:

Intercurrent Event	**Strategy for Addressing Intercurrent Events and Its Description**
Treatment discontinuation	Treatment policy approach will be used, i.e., a subject's event will be used in the analysis regardless of treatment disposition.

Subsequent anti-cancer therapy	Treatment policy approach will be used, where the occurrence of the intercurrent event is irrelevant. Thus, a subject's event is used in the analysis regardless of starting subsequent anti-cancer therapy.

5.3.3. Analysis Methods

Cohort 1 and Cohort 2 will be assessed independently, i.e., as if there were 2 separate studies. The primary efficacy analysis will be based on the ITT Analysis set that includes all randomized subjects. The survival curves of OS will be described using KM methods together with estimated median (in months), etc., with 95% CI for each treatment group. A stratified log-rank test will be used to compare the treatment groups (within each cohort) at an overall alpha level of 0.05 for the primary analyses. The stratification factors to be used in the analysis are as follows: ECOG performance status (0 or 1 vs. 2), disease distribution (presence vs. absence of visceral [lung, liver, or bone] metastases) and region (North America vs. EU vs. ROW). If necessary, data will be pooled for these stratification factors based on the rules defined in Section 5.1.2.

5.3.3.1. Additional analysis of Primary Endpoint OS

Additional sensitivity and exploratory analyses will be performed on the primary endpoint (OS). One sensitivity analysis of OS will compare the 2 treatment arms using log-rank test without adjusting for stratification factors. Other analyses are specified in the following subsections.

5.3.3.1.1. Covariate adjusted analysis

Additional exploratory analyses on OS will be performed using a selected set of potential prognostic variables (obtained at or before baseline) as covariates in Cox regression models. Potential prognostic factors are, but not limited to:

- Age (< 65 vs. ≥ 65)
- Gender
- Race: (Caucasian vs. non-Caucasian)
- Region (NA vs. EU vs. ROW)
- Baseline ECOG (0-1 vs. 2)
- Baseline disease distribution (presence vs. absence of visceral metastases in lung, liver or bone)

Each factor will be assessed individually for prognostic value ($p < 0.05$) using univariate Cox model. Factors that are deemed to have prognostic value will be included as covariates in a multivariate Cox model to assess their significance in the presence of other factors. Selection methods will be used to identify the final set of prognostic factors (p-value < 0.10). Treatment will then be added to this final model to assess the effect of treatment when adjusted for these prognostic factors.

5.3.3.1.2. Subsequent Anti-Cancer Therapy and Crossover Adjusted Analysis

Additional exploratory analysis may be performed to assess the impact of subsequent anti-cancer therapy on overall survival as appropriate, such as censoring OS at the start of subsequent anti-cancer therapy and using Inverse Probability of Censoring Weighting (IPCW) (Robins et al., 2000) method to adjust for subsequent anti-cancer therapy.

Treatment crossover after disease progression may occur and its impact on OS may be explored.

5.3.3.1.3. Subgroup Analysis

Table 5 of section 5.6.4 includes a list of subgroups and their definitions.

Subgroup analyses will be conducted to assess efficacy consistency across patient populations with different demographics or baseline characteristics. A forest plot will be used to summarize graphically the subgroup analysis results. Kaplan-Meier curves will be presented for each subgroup variable to provide additional supportive evidence.

5.4. Secondary Endpoints Analysis

5.4.1. Key Secondary Endpoints

5.4.1.1. Progression- Free Survival (PFS)

The PFS is defined as duration (in days) from the date of randomization to the date of disease progression (or relapse from CR) assessed per RECIST v1.1 by the investigator or death, whichever is reported first. PFS will be censored at the date of the last adequate disease assessment for subjects who do not have disease progression and are alive, as well as for subjects with unknown disease progression or unknown survival status as of the clinical cutoff date. Also, if there is no post-baseline tumor assessment for a subject, PFS will be censored on the date of randomization.

Similar to the procedure for the primary analysis of OS, PFS will be analyzed by treatment group using the method specified in Section 5.1.

The details regarding the handling of missing assessment and censoring for PFS analysis are presented in Table 1.

Table 1.: Censoring rules for PFS Analysis

Situation	Date of Progression or Censoring	Outcome
Disease progression prior to the start of anti-cancer therapy	Date of disease progression	PFS event
Disease progression after the start of anti-cancer therapy	Date of last adequate assessment*	Censored
Death without any documented disease progression	Date of death	PFS event
No baseline and/or no post-baseline assessment, no subsequent anti-cancer therapy after study treatment, no death	Date of Randomization	Censored
Other: • Lost to follow-up, withdraw consent before any documented death or disease progression	Date of last adequate assessment*	Censored

- No documented disease progression at data cutoff

*Adequate disease assessment is defined as there is sufficient data to evaluate a subject's disease status.

5.4.1.2. Objective Response Rate (ORR)

The ORR is defined as the proportion of subjects who achieve CR or PR, as assessed per RECIST v1.1 by the investigator among ITT Analysis set of a treatment group.

Subjects will be considered as non-responders if they do not have CR or PR while on study, or do not have a baseline or post-baseline tumor assessment, or do not have adequate baseline tumor evaluation, or die, or have progressive disease, or drop out for any reason or take subsequent therapy prior to reaching a CR or PR.

A stratified Cochran-Mantel-Haenszel (CMH) test will be used to test treatment difference at a significance level of 0.05 and to estimate the relative risk and p-values between the treatment groups. The 95% confidence limits (CLs) for the relative risks will also be calculated based on the Wald statistics. The stratification factors to be used in the analysis are as follows: region (North America vs. EU vs. ROW), ECOG performance status (0 or 1 vs. 2), and disease distribution (presence vs. absence of visceral metastases in lung, liver, or bone). If necessary, data will pooled for the stratification factors based on the rules defined in Section 5.1.2 with number of events replaced by number of patients. Additional subgroup analysis will be performed when appropriate.

5.4.1.3. PRO-Related Endpoints

Similar to the procedure for the primary analysis of OS, TUSD-3 will be analyzed by treatment group using the method specified in Section 5.1. Additional details regarding the derivation and analysis of TUSD-3 as well as other PRO endpoints can be found in Section 5.6.2 and in a separate psychometric analysis plan (PAP).

5.4.2. Supportive Secondary Endpoint

5.4.2.1. Duration of Response (DOR)

For subjects achieve an overall response of CR or PR only (responders), the DOR is defined as duration (in days) from the date of initial documentation of an overall response of CR or PR to the date of first documented evidence of progressive disease (or relapse for subjects who experience CR during the study) or death. The DOR will only be calculated for the subgroup of subjects with an overall response of CR or PR. The censoring for DOR is similar to the censoring rules for PFS.

DOR will be summarized by treatment group (erdafitinib vs. chemotherapy or pembrolizumab) using the KM method and displayed graphically when appropriate. The distribution of DOR will be described using KM methods, reporting estimated median (in months) with 95% CI for each treatment group. Overall treatment group estimates will be provided.

5.5. Safety Analyses

Safety assessments will be based on medical review of AE reports and the results of vital sign measurements, electrocardiograms, physical examinations, clinical laboratory tests, ECOG performance status, ophthalmologic examinations, Amsler grid test, and other safety evaluations at specified time points as described in the Time and Events Schedule of the study protocol. Safety data will be analyzed using the Safety Analysis Set. The baseline value for safety assessment is defined as the value collected at the time closest to, but prior to, the administration of the first dose of study drug.

In general, in instances where there are multiple records at a given visit date for lab parameters associated with disease assessment, select the latest lab value as the unique lab value for analysis.

5.5.1. Extent of Exposure

The number and percentage of subjects who receive study agent (erdafitinib, vinflunine, docetaxel, or pembrolizumab) at each dose level will be summarized by treatment group. The administration of study agents will be presented, by medication administered within each treatment group and will be described in terms of the total number of cycles administered, the median of cycles administered, dose intensity, dose modifications (including dose reduction and dose interruption),and dose up-titration (for erdafitinib only).

For daily dosing, treatment duration for the study will be calculated as:

date of last dose of study drug – date of first dose of study drug + 1.

For non-daily dosing, if subjects died before end date of last cycle (incomplete last cycle), treatment duration for the study will be calculated as:

death date – date of first dose of study drug + 1.

Subject-years of exposure are calculated as days of exposure/365.25. Subject-years will be presented by treatment group.

Descriptive statistics for treatment duration (N, mean, SD, median, and range [minimum, maximum]) will be presented by treatment group using the safety population.

Duration of treatment will be summarized in the following duration categories: [<6 weeks, 6-<12 weeks, 12-<18 weeks, 18-<24 weeks, 24-<30 weeks, 30-<36 weeks, 36-<42 weeks, 42-<48 weeks, 48-<54, 54-<60, 60-<66, 66-<72 weeks, and additional duration categories as needed] by treatment group.

Total dosing days are defined as the total number of days that study agent (erdafitinib) has been administered to the subject (excluding days of study agent interruption).

The number (%) of subjects with a dose adjustment will be summarized by treatment group and cycle. Reasons for dose adjustments will also be summarized.

Descriptive statistics will be presented for study agent (erdafitinib) using the following parameters:

- Number of study agent administrations
- Cumulative total dose
- Mean daily dose
- Relative dose intensity

The mean daily dose of study agent is calculated as (sum of total daily dose during the treatment phase)/treatment duration.

Relative dose intensity is defined as cumulative total dose/ cumulative planned total dose which is based on initial planned dose displayed as percentages.

For subjects treated with vinflunine, docetaxel, or pembrolizumab, descriptive statistics will be presented for using the following parameters:

- Number of study agent administrations
- Cumulative total dose
- Mean cycle dose
- Relative dose intensity

The mean cycle dose is calculated as (sum of total dose during the treatment phase)/number of cycles.

Relative dose intensity is defined as cumulative total dose/cumulative planned total dose which is based on initial planned dose displayed as percentages.

A by-subject listing will present all the study agents that have been taken by the subject, which include cohort, treatment group, study day, cycle day, name of study agent and doses. Any dose adjustment/withhold/delays and the reasons need to be presented accordingly.

Study intervention compliance will be summarized descriptively. See Appendix 5 (Section 6.5) for further details.

5.5.2. Adverse Events

The verbatim terms used in the case report form (CRF) by investigators to identify AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The severity of all AEs will be graded according to NCI-CTCAE Version 4.03.

5.5.2.1. Treatment-Emergent Adverse Events

The treatment-emergent period is defined as the time from first dose date through 30 days after last dose date, or day before subsequent anti-cancer therapy, whichever occurs first; or any AE that is considered treatment-related (very likely, probably, or possibly related) regardless of the

start date of the event; or any AE that is present at baseline but worsens in toxicity grade or is subsequently considered drug-related by the investigator. Treatment-related AEs are those assessed by investigator as being possible, probable or very likely related to study drug.

Unless otherwise specified, at each level (e.g., system organ class [SOC] and/or preferred term [PT]) of subject summarization in reporting the incidence of the AE, a subject is counted once if one or more events were recorded.

The following TEAEs will be summarized by treatment group, which is also listed in Table 2:

- Overall summary of the incidence of TEAEs, including
    - Any TEAEs
    - TEAEs by toxicity grade
    - TEAEs by relationship to study agent,
    - Serious TEAEs
    - Treatment-related AEs
    - Grade 3 or worse TEAEs
    - TEAEs leading to treatment discontinuation
    - TEAEs leading to death
    - TEAEs leading to dose modification
    - TEAEs of clinical interests
    - TEAEs of special interests
    - TEAE with fatal outcomes within 30 days of last dose
- TEAEs by SOC, PT, toxicity grade, and relationship, with incidence frequency higher than 5%
- TEAEs by PT, toxicity grade, and relationship, with incidence frequency higher than 5%
- Serious TEAEs by SOC, PT, toxicity grade, and relationship, with incidence higher than 2%
- Serious TEAEs by PT, toxicity grade, and relationship, with incidence frequency higher than 2%
- Grade 3 or worse TEAEs by PT, toxicity grade, and relationship, with incidence frequency higher than 2%
- TEAEs leading to treatment discontinuation by PT, toxicity grade, and relationship
- TEAEs leading to death by PT and toxicity grade
- TEAEs leading to dose modification by PT and toxicity grade
- TEAEs of clinical interests by PT and toxicity grade
- Other safety observations by PT and toxicity grade
- TEAE with fatal outcome within 30 days of last dose by reason of death

In addition, overall summary of TEAEs as well as TEAEs by SOC, PT and toxicity grade may be generated for selected subgroups as defined in Table 5 in Section 5.6.4.

Table 2. Summary of Treatment-Emergent Adverse Event Analyses to be Performed

Category	Analysis	Sorted By	Cut off	Treatment-Related TEAE
General	Overall summary			✔
	TEAEs	SOC+ PT+ toxicity grade; PT+ toxicity grade	5%	✔
	Serious TEAEs	SOC+ PT + toxicity grade; PT+ toxicity grade	2%	✔
	Grade 3 or worse TEAE	PT+ toxicity grade	2%	✔
	TEAEs leading to treatment discontinuation	PT + toxicity grade		✔
	TEAEs leading to death	PT + toxicity grade		
	TEAEs leading to dose modification or modification	PT + toxicity grade		
	TEAEs of clinical interest (Hemorrhagic events)	PT + toxicity grade		
	Other safety observations (e.g. other malignancies, eye disorder)	PT + toxicity grade		
	Deaths within 30 days of last dose	Reason for death		
Selected subgroup	Overall summary			
	TEAEs	SOC+ PT+ toxicity grade		

5.5.2.2. Deaths

Deaths will be displayed by actual treatment received. Frequencies for the following parameters will be included in the summary table:

- Number of subjects who died
- Cause of death

A listing of subjects who died will be provided.

5.5.3. Additional Safety Assessments

5.5.3.1. Clinical Laboratory Tests

All clinical laboratory tests will be displayed for the subjects included in the safety analysis set. If a subject has repeated laboratory values for a given time point, the worst value will be used.

Descriptive statistics will be presented for all chemistry and hematology laboratory tests at scheduled time points.

Change from baseline to each time point will be summarized for chemistry and hematology by treatment group. Unscheduled laboratory test results will be listed and also included in the laboratory shift tables.

Number and percentage of subjects with post-baseline clinically important laboratory values and/or markedly abnormal post-baseline values will be presented by treatment group.

The clinically important laboratory findings to be reported are described below:

- AST (U/L): ≥2x ULN
- ALT (U/L): ≥2x ULN
- Alkaline phosphatase (U/L): High (≥3xULN)
- Bilirubin (total) ≥2x ULN

Markedly abnormal laboratory findings to be reported are described below:

- AST (U/L) or ALT(U/L): ≥3x ULN
- AST(U/L) or ALT (U/L): ≥5x ULN
- Grade 4 NCI-CTCAE

Applicable laboratory results will be graded according to NCI-CTCAE version 4.03.

A listing for all subjects with clinically important laboratory values will be provided and a summary table of the number of such subjects will be provided by treatment group.

Descriptive statistics of clinical laboratory results and their changes from baseline will be presented by treatment group.

Shift tables will be provided summarizing the shift in laboratory values from baseline to the End-Of-Treatment visit with respect to abnormality criteria (low, normal, high).

Shift summaries from baseline laboratory value to the worst on-treatment US NCI-CTCAE, version 4.03 grade in chemistry and hematology tests with US NCI-CTCAE, version 4.03 will be presented.

5.5.3.2. Vital Signs and Physical Examination Findings

Vital sign and physical examination summaries and reporting of screening will be based on the safety analysis set. Continuous vital sign parameters include weight, height, and pulse, will be summarized by treatment group and scheduled visits.

5.5.3.3. Electrocardiogram

The ECG parameters that will be summarized over-time are heart rate, RR interval, PR interval, QRS interval, QT interval, and QTc using the following correction methods: Fridericia's formula (QTcF). QTcF (Fridericia) will be used for assessment of QTc interval.

Fridericia's formula: QTcF (msec) = QT (msec) * $(HR(bpm)/60)^{1/3}$

Values outside the normal range will be flagged as follows.

Observed:

- Heart rate: L <50 bpm; H > 100 bpm
- RR interval: L < 600 ms; H > 1000 ms
- QT interval: H > 500 ms
- QTc interval: H > (450 ms for males, 470 ms for females); increase to >500 ms

Change from baseline:

- QTc: 30-60 ms increase; increase >60 ms

All treatment-emergent abnormal findings will be tabulated, displaying the number of subjects with abnormal findings after dosing. An abnormal finding is considered to be treatment emergent if it occurred during treatment and up to 30 days after the last dose.

Descriptive statistics for the actual values and changes from baseline in ECGs will be tabulated by time point based on safety population. No statistical testing will be performed. If more than 1 ECG

measurements are repeated at a visit, they will be averaged. The averaged value will be considered the 'Visit' ECG results.

Maximum QTc intervals and maximum changes from baseline will also be summarized similarly in a separate display.

The interpretation of the ECGs and percentage of subjects meeting the normality criteria will be summarized over time by treatment group in each cohort.

5.5.3.4. Eastern Cooperative Oncology Group (ECOG) Performance Status

Eastern Cooperative Oncology Group (ECOG) performance status will be summarized at each scheduled time point. Descriptive statistics will be calculated at baseline and for observed values and change from baseline at each scheduled time point by treatment group in each cohort.

5.5.3.5. Ophthalmologic Examinations and Amsler Grid Test

Summary of ophthalmologic examinations and Amsler grid test may be provided, including shift tables. Listings of ophthalmologic examinations and Amsler grid test will be generated.

5.5.3.6. Other Safety Parameters

Pregnancy testing results will be presented in a by-subject listing.

5.6. Other Analyses

5.6.1. Pharmacokinetics

Concentration data listing and summary statistics as described below will be performed on the Pharmacokinetic-evaluable Analysis Set, which includes all randomized subjects who received at least 1 dose of -erdafitinib and had at least 1 evaluable pharmacokinetic sample obtained posttreatment. Key characteristics of the Pharmacokinetic-evaluable Analysis Set will be summarized. Erdafitinib plasma concentration data will be listed for all subjects with available evaluable pharmacokinetic sample. Concentration data points will be excluded from the listing based on lack of accurate assessment on critical aspects of concentration data (eg, missing information of dosing and sampling times). All concentrations below the lowest quantifiable concentration or missing data will be labeled as such in the concentration database. All subjects and samples excluded from the listing will be clearly documented in the study report.

Descriptive statistics (N, mean, standard deviation, range, coefficient of variation (%CV), and IQ range) will be used to summarize plasma concentrations at each nominal sampling time point. Graphical exploration of data may be performed as deemed useful. Measurements of erdafitinib free concentration and free fraction may also be listed and summarized, if available.

Plasma erdafitinib concentrations below the LLOQ will be imputed as zero in the summary statistics. All subjects and samples excluded from the analysis will be clearly documented and presented in a listing.

5.6.2. Patient Reported Outcomes and Health Economics

5.6.2.1. Patient-reported Outcomes (PROs)

An important aspect of patient-centered drug development relates to describing the treatment benefit using the patient's perspective through use of Patient-Reported-Outcomes (PRO). For this purpose, three PRO instruments have been implemented in this study: the FACT-Bl, the PGIS, and the EQ-5D-5L. These instruments allow capturing symptom severity for symptoms experienced by patients with bladder cancer (including key urinary bladder cancer symptoms), impact of disease and treatment on different aspects of patients' well-being and overall synthetic health status or Health-related Quality of Life (HRQOL). These PRO data will be collected electronically (ePRO) according to the Time and Events Schedule to assess change over time and difference between treatment groups in each cohort. EQ-5D-5L will also be collected in the Follow-up Phase until start of subsequent anti-cancer therapy.

5.6.2.1.1. FACT-Bl

The FACT-Bl consists of 39 items, with 5-point Likert scales, covering 5 primary domains: physical well-being, social/family well-being, emotional well-being, functional well-being and additional concerns for patients with bladder cancer. The response options range from "Not at all" to "very much." In addition to the scores defined by the author, a Urinary bladder cancer Symptoms (US) score will be derived to assess the severity of urinary symptoms over time (see Section 5.6.2.2.3).

Other FACT-Bl scores will be derived following the scoring algorithm developed by the authors of the instrument and presented in Appendix 7 (Section 6.7).

5.6.2.1.2. PGIS

The PGIS is a single question regarding the patient report of disease severity: "Considering all aspects of your bladder cancer symptoms right now would you say your bladder cancer symptoms are absent (0), mild(1), moderate(2), severe(3), or very severe(4)?". The PGIS is an anchor question that will be used to establish the magnitude of meaningful change thresholds for each PRO derived score allowing to classify patients as improved / stable / worsened.

5.6.2.1.3. EQ-5D-5L

The EQ-5D-5L is a generic measure of health status. For purposes of this study, the EQ-5D-5L will be used to generate utility scores for use in cost effectiveness analyses. The EQ-5D-5L is a 5-item questionnaire that assesses 5 domains including mobility, self-care, usual activities, pain/discomfort and anxiety/depression plus a visual analog scale (VAS) rating "health today" ranging from 0 (worst imaginable health state) to 100 (best imaginable health state) (Herdman et al., 2011). The scores for the 5 dimensions are used to compute a single utility score representing the general health status of the subject. The EQ-5D-5L asks respondents to select their response based on their current health and takes less than 5 minutes to complete.

5.6.2.2. Analysis Methods

The ITT Analysis set will be used for the PRO analyses. All analyses are done by treatment group in each cohort unless it is specified. Based on the PRO assessment date, the associated visit will be mapped and used for analysis. For subjects with multiple records at the same visit, the closest one to the visit date will be selected as scheduled assessment, and others will be unscheduled assessments. For those with multiple records on the same assessment date, the first one by time point and sequence number will be selected as scheduled assessment. Only scheduled visits will be presented in the by-visit analysis. All assessments will be included in the time-to-event analysis. Analysis may be stratified by stratification factors, prior anti PD-(L)1 (yes vs. no), region (North America vs. EU vs. ROW), ECOG performance status (0 or 1 vs. 2), and disease distribution (presence vs. absence of visceral metastases), when appropriate. For subjects receiving subsequent anti-cancer therapies after discontinuing study drug, only PRO measurements before the initiation of subsequent therapy will be included in analysis unless otherwise stated.

5.6.2.2.1. Compliance with PRO Assessments

Missing PRO assessments are calculated as the expected number of assessments for a particular visit minus the actual number of assessments received for that visit. Compliance (% received and % missing forms) will be tabulated by treatment group and overall for baseline, each scheduled visit in each cohort. Expected number of assessments per visit will be determined by subject-level study completion status. These compliance tables for each cohort will follow the table below.

Table 3.: Compliance of PRO assessments

Timing of Assessment	Treatment Arm 1 (N = xxx)			Treatment Arm 2 (N = xxx)		
	Number of Forms, n (%)			Number of Forms, n (%)		
	Expected	Received	Missing	Expected	Received	Missing
Baseline	xxx	xx (xx.x)	xx (xx.x)	xxx	xx (xx.x)	xx (xx.x)
C1D1	xxx	xx (xx.x)	xx (xx.x)	xxx	xx (xx.x)	xx (xx.x)
C1D14	xxx	xx (xx.x)	xx (xx.x)	xxx	xx (xx.x)	xx (xx.x)
…	xxx	xx (xx.x)	xx (xx.x)	xxx	xx (xx.x)	xx (xx.x)
Total	xxx	xx (xx.x)	xx (xx.x)	xxx	xx (xx.x)	xx (xx.x)

5.6.2.2.2. Change Scores From Baseline Over Time

Descriptive statistics (number of observations, mean, standard deviation, minimum, maximum) of scores at baseline and change from baseline at each scheduled visit will be produced for FACT-BI and EQ-5D-5L scores. For these descriptive statistics, the scores obtained at the end of study treatment visit will be assigned to a cycle number as the next cycle that would have been completed if the subject were to have continued receiving study treatment. EQ-5D-5L scores, separate summaries will be performed at treatment phase and follow-up phase. For PGIS, frequency count and percentage of each score level (1 to 5) over time will be provided.

A mixed effects model with repeated measures analysis will be conducted estimating change from baseline at each scheduled visit between two treatment arms. Subjects in the ITT Analysis set with a baseline value and at least one post-baseline value are included in the analysis.

Change from baseline will be fitted to a mixed effects model including subjects as a random effect, and baseline value, treatment group, time in month, treatment-by-time interaction, and stratification factors as fixed effects. Plots of the observed and model-estimated mean change from baseline with standard deviation and standard error, respectively over time by treatment group will be displayed.

5.6.2.2.3. Urinary Bladder Cancer Symptom Score and Measurement Properties

The Urinary bladder cancer Symptoms (US) score will be obtained by using scores of the following items: BL1 – have trouble controlling my urine, BL2 – I urinate more frequently than usual, BL3 – It burns when I urinate. Content validity of the US score will be based on the result of the qualitative study (Protocol Number: JN1027C v0.1). The method to create US score will be guided by the qualitative study. In addition to the US score described in the SAP, additional details is also provided in a separate PRO PAP.

5.6.2.2.4. Derivation of Meaningful Change thresholds

Methods for deriving meaningful change thresholds for the PROs are detailed in a separate PRO PAP.

5.6.2.2.5. Time to Urinary Bladder Cancer Symptoms Deterioration

Time to Urinary bladder-cancer Symptoms Deterioration (TUSD)-3 will be generated based on an US score from 3 items from the FACT-Bl (BL1. Urinary incontinence, BL2. Urinary frequency, BL3. Urinary pain) and is defined as the first time to increase in US score from the day of randomization beyond a meaningful change threshold compared to baseline. For subjects whose US score did not worsen to the magnitude of the meaningful change threshold from baseline before disease progression, death or dropout of study, TUSD-3 will be censored at the last post-baseline FACT-Bl assessment, or the randomization date if there is no post-baseline FACT-Bl assessment. The censoring rules are listed in Table 4.

Each endpoint, TUSD-BL1, TUSD-BL2, and TUSD-BL3, will be derived based on time to deterioration from baseline in each urinary bladder-cancer symptom (BL1. Urinary incontinence, BL2. Urinary frequency, BL3. Urinary pain) and is defined as the time to increase in respective urinary bladder-cancer symptom from the day of randomization beyond a meaningful change threshold compared to baseline. For subjects whose respective US score did not worsen to the magnitude of the meaningful change threshold from baseline before disease progression, death or dropout of study, each endpoint will be censored at the last post-baseline FACT-Bl assessment, or the randomization date if there is no post-baseline FACT-Bl assessment. Censoring rules are listed in Table 4.

For each endpoint, the median time to deterioration will be estimated using a Kaplan-Meier method. Additionally, hazard ratio (Erdafitinib /SoC) and associated 95% confidence interval will be estimated by stratified Cox's proportional hazards model with stratification variables. In cases where median values cannot be computed because less than 50% of subjects experienced deterioration, 25th percentiles will be reported and compared instead.

Table 4.: Censoring Scheme for Urinary Bladder Cancer Symptoms Deterioration*

Situation	Date of Progression or Censoring	Outcome
No baseline and/or no post-baseline assessment	Date of Randomization	Censored
Deterioration before death or disease progression or subsequent anti-cancer therapy occurs	Date of the FACT-Bl assessment	Deterioration
Deterioration after missing more than one consecutive scheduled FACT-Bl assessment	Date of the last FACT-Bl assessment before missing visits	Censored
Deterioration after subsequent anti-cancer therapy	Date of the last FACT-Bl assessment before subsequent anti-cancer therapy	Censored
Lost to follow-up, withdraw consent, end of study before death or disease progression or Deterioration occurs	Date of the last FACT-Bl assessment	Censored
Disease Progression before Deterioration or death occurs	Date of the last FACT-Bl assessment before disease progression	Censored
Death before Deterioration or Disease Progression occurs	Date of the last FACT-Bl assessment	Censored

* Urinary Bladder Cancer Symptoms Deterioration: US score is higher than the magnitude of the meaningful change threshold

5.6.3. Concordance of Central Laboratory and Local Historical test results of Selected FGFR Aberrations

To assess the concordance of the central laboratory and the local historical test results of selected FGFR aberrations, a table summarizing the percent agreement between local and central testing for each FGFR alteration/variant type will be provided.

5.6.4. Definition of Subgroups

Subgroup analyses will be performed as appropriate to evaluate the consistency in the selected efficacy and safety endpoints. Table 5 provides the categorical variables that will be used for subgroup analysis.

Table 5.: Categorical variables used for subgroup analysis.

Subgroup	Variant	Definition
Region†	1	North America (NA) Europe (EU) Rest-of-the-World (ROW)
Baseline ECOG performance status	1	0 – 1 2
Disease distribution (presence or absence of visceral metastases: lung, liver or bone)	1	Presence Absence
Bone Metastasis	1	Yes No
Liver Metastasis	1	Yes No
Lung Metastasis	1	Yes No
Primary tumor location	1	Lower tract - Bladder - Urethra - Prostatic urethra Upper tract - Renal pelvis - Ureter
FGFR alteration type	1	Translocations Mutations
PD-L1 status	1	Positive (Combined Positive Score (CPS) ≥ 10) Negative (CPS < 10)
Baseline creatinine clearance	1	< 30 mL/min 30 – <60 mL/min ≥ 60 mL/min
Baseline hemoglobin level	1	<10 g/dL ≥ 10 g/dL
Gender	1	Female Male
Age group	1	< 65 years ≥ 65 years
Race‡	1	American Indian or Alaska Native Asian Black or African American Native Hawaiian or other Pacific Islander White Not reported

Ethnicity	1	Hispanic/Latino Not Hispanic/Latino Not Reported

† North America (NA) includes USA and Canada; Europe (EU) includes Austria, Belgium, Czech Republic, Denmark, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Poland, Portugal, Russia, Slovakia, Spain, Sweden, Switzerland, Turkey, Ukraine, United Kingdom and any other countries/territories belonging to geographical Europe. ROW includes Australia, China, India, Japan, South Korea, Taiwan, Thailand, Argentina, Brazil, Mexico and any other countries/territories not included in the NA or EU.
‡ If number of subjects from 'American Indian or Alaska Native', 'Asian', 'Native Hawaiian or other Pacific Islander', and 'Not reported' are less than 10 for a treatment group, then combined these subjects into a new subgroup named 'Other.'

5.7. Interim Analysis

There is one planned interim analysis for each cohort. The interim analyses are planned after an approximately 65% information fraction (approximately 136 of 208 deaths) in Cohort 1 and (approximately 172 of 264 deaths) in Cohort 2 will have occurred. Both superiority and futility will be assessed for Cohort 1 and Cohort 2. The classical O'Brien-Fleming boundaries will be used for these assessments. The stopping boundaries will be implemented by Lan-DeMets spending function to control the Type 1 error at the 0.05 significance level overall for each cohort. The futility assessment for each cohort is not binding. Also, each cohort may be stopped when the HR is 1.0 or greater, taking into consideration the totality of the data. Specific details will be provided in the interim analysis plan (IAP).

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

AE	adverse event
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BUN	blood urea nitrogen
CI	confidence interval
CL	confidence limit
CPS	combined positive score
CR	complete response
CRF	case report form
CMH	Cochran-Mantel-Haenszel
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DMC	Data Monitoring Committee
DOR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EQ-5D-5L	European Quality of Life – 5 Dimensions-5 Levels
FACT-BL	Functional Assessment of Cancer Therapy – Bladder
HR	hazard ratio
IA	interim analysis
IAP	interim analysis plan
ICH	International Conference on Harmonization
IDMC	Independent Data Monitoring Committee
IHC	immunohistochemistry
IPCW	Inverse Probability of Censoring Weighting

ITT	intent-to-treat
KM	Kaplan-Meier
LLOQ	lower limit of quantification
LOCF	last observation carried forward
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	objective response rate
OS	overall survival
PD-(L)1	programmed death-ligand 1
PDS	protocol deviation specification
PFS	progression-free survival
PGIS	Patient-Global Impression of Severity
PK	pharmacokinetic(s)
PR	partial response
PRO	Patient-reported Outcome
PT	preferred term
RECIST	Response Evaluation Criteria in Solid Tumors
ROW	Rest-of-the-World
SAP	statistical analysis plan
SD	standard deviation
SES	standardized effect size
SSG	Statistical Support Group
SOC	system organ class
TCR	T-cell receptor
TEAE	treatment-emergent adverse event
TUSD	time to urinary symptom deterioration
WHO	World Health Organization
WHO-DD	World Health Organization Drug Dictionary

6.2. Appendix 2 Demographics and Baseline Characteristics

The number of subjects in each analysis set will be summarized and listed by treatment group. In addition, the distribution of subjects by region (North America, EU and ROW) will be presented unless otherwise noted.

Subject demographics and baseline disease characteristics will be summarized using descriptive statistics, which include the following items:

Table 6 presents a list of the demographic variables that will be summarized by treatment group for the safety analysis set.

Table 6: Demographic Variables

Continuous Variables:	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum])
Weight (kg)	
Height (cm)	
Categorical Variables	
Age (<65 years, 65 - 69 years, 70 - 74 years, ≥75 years)	Frequency distribution with the number and percentage of participants in each category.
Sex (male, female)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple Not reported)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino, Not reported)	

[a]If multiple race categories are indicated, the Race is recorded as 'Multiple'

Table 7 presents a list of the baseline disease characteristics variables that will be summarized by treatment group for the safety analysis set.

Table 7: Baseline Characteristics

Continuous Variables:	**Summary Type**
Time from date of initial diagnosis to 1st dose (months)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum])
Time from progression/relapse on the last line of prior chemotherapy to 1st dose (months)	
Hematology: hemoglobin, platelet count, white blood cell (WBC) count, absolute neutrophil count (ANC)	
Chemistry: alanine aminotransferase (ALT), chloride, albumin, creatinine, creatinine clearance, alkaline phosphatase, magnesium, aspartate aminotransferase (AST), alpha-1-acid glycoprotein, bicarbonate, phosphate, blood glucose, potassium, blood urea nitrogen (BUN), sodium, total bilirubin, total protein, calcium, parathyroid hormone (PTH)	
Categorical Variables	
Geographic region (North America, Europe, Rest of the World)	Frequency distribution with the number and percentage of participants in each category.
Primary tumor location (Lower tract: Bladder, Urethra, Prostate; Upper tract: Renal, Pelvis, Ureter)	
Type of histology (Transitional Cell Carcinoma, Transitional Cell Carcinoma with minor components (<50% overall) of variant histology)	
Tumor at initial diagnosis (TX, T0, Ta, Tis, T1, T2, T3, T4)	
Lymph node at initial diagnosis (NX, N0, N1, N2, N3)	
Metastatic at initial diagnosis (N, M0, M1)	
Bladder Cancer Stage at initial diagnosis (0a, 0is, I, II, III, IV)	
Baseline ECOG (0, 1, 2)	
PD-L1 status (positive, negative)	
Number of Prior Lines of Systemic Therapy (1, 2)	
Disease distribution at study entry (Presence of visceral metastases: Lung, Liver, Bone; Absence of visceral metastases (lung, liver, or bone); Not reported)	
FGFR alteration type: Specific Translocation (FGFR3:BAIAP2L1, FGFR2:BICC1, FGFR2:CASP7, FGFR3:TACC3v1, FGFR3:TACC3v3), Specific Mutation (FGFR3 S249C, R248C, G370C, Y373C)	

6.3. Appendix 3 Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact subjects' safety or well-being, or the integrity and/or result of the clinical study. Subjects with major protocol deviations need to be identified prior to database lock and summarized by the following categories:

- Entered the study but did not satisfy criteria
- Developed withdrawal criteria but not withdrawn
- Received wrong treatment or incorrect dose
- Received a disallowed concomitant treatment
- Others

A Protocol Deviation Specification (PDS) has been developed to provide more information about the major protocol deviations. Periodic meetings are required to investigate each potential protocol deviation.

6.4. Appendix 4 Prior and Concomitant Medications

Prior and Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of first dose (partial or complete) of study agent. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study agent, including those that started before and continued after the first dose of study agent.

The number and percentage of subjects taking concomitant medications from the first dose through the end of treatment phase (up to 30 days after the last dose of study drug) will be tabulated by the Anatomical Therapeutic Chemical (ATC) classification pharmacological subgroup and WHO drug generic term for each treatment group in the safety population. The proportion of subjects who receive each concomitant medication will be summarized as well as the proportion of subjects who receive at least one concomitant medication. A by-subject listing will also be presented for concomitant medication. Prior medications will be summarized by treatment group and ATC terms.

6.5. Appendix 5 Intervention Compliance

Study agent compliance will be summarized descriptively based on the safety analysis population.

- For subjects on erdafitinib (Arm 1A and 2A), study agent compliance will be calculated as follows:

  Study agent compliance (%) = (actual dose taken/total dose prescribed) x100.

- For subjects on chemotherapy (vinflunine or docetaxel, Arm 1B) or pembrolizumab (Arm 1B and 2B), study agent compliance will be calculated as follows:

  Study agent compliance (%) = (actual IV dose infused /total IV dose prescribed) x100.

6.6. Appendix 6 Adverse Events of Special Interest and Clinical Interest

Adverse events of special interest and clinical interest are defined as follows:

AE Special Interest Category	Preferred Term
Central Serous Retinopathy	Retinal detachment
Central Serous Retinopathy	Vitreous detachment
Central Serous Retinopathy	Retinal oedema
Central Serous Retinopathy	Retinopathy
Central Serous Retinopathy	Chorioretinopathy
Central Serous Retinopathy	Detachment of retinal pigment epithelium
Central Serous Retinopathy	Detachment of macular retinal pigment epithelium
Central Serous Retinopathy	Macular detachment
Central Serous Retinopathy	Serous retinal detachment
Central Serous Retinopathy	Subretinal fluid
Central Serous Retinopathy	Retinal thickening
Central Serous Retinopathy	Chorioretinitis
Central Serous Retinopathy	Serous retinopathy
Central Serous Retinopathy	Maculopathy
Central Serous Retinopathy	Choroidal effusion
Central Serous Retinopathy	Central serous chorioretinopathy

AE Clinical Interest Category	Preferred Term
Eye Toxicity	Retinal thickening
Eye Toxicity	Blepharitis
Eye Toxicity	Cataract
Eye Toxicity	Cataract Subcapsular
Eye Toxicity	Conjunctival haemorrhage
Eye Toxicity	Conjunctival hyperaemia
Eye Toxicity	Conjunctival hyperemia
Eye Toxicity	Conjunctival Irritation
Eye Toxicity	Corneal erosion
Eye Toxicity	Corneal infiltrates
Eye Toxicity	Dry eye
Eye Toxicity	Eye inflammation
Eye Toxicity	Eye irritation
Eye Toxicity	Eye pain
Eye Toxicity	Foreign body sensation in eyes
Eye Toxicity	Keratitis
Eye Toxicity	Lacrimation increased
Eye Toxicity	Night blindness
Eye Toxicity	Ocular hyperaemia
Eye Toxicity	Photophobia
Eye Toxicity	Vision blurred
Eye Toxicity	Visual acuity reduced
Eye Toxicity	Visual impairment
Eye Toxicity	Xanthopsia
Eye Toxicity	Xerophthalmia
Eye Toxicity	Chorioretinitis
Eye Toxicity	Conjunctivitis
Eye Toxicity	Ulcerative keratitis
Eye Toxicity	Retinal detachment
Eye Toxicity	Vitreous detachment
Eye Toxicity	Retinal oedema

Eye Toxicity	Retinopathy
Eye Toxicity	Chorioretinopathy
Eye Toxicity	Detachment of retinal pigment epithelium
Eye Toxicity	Detachment of macular retinal pigment epithelium
Eye Toxicity	Macular detachment
Eye Toxicity	Serous retinal detachment
Eye Toxicity	Subretinal fluid
Eye Toxicity	Retinal thickening
Eye Toxicity	Chorioretinitis
Hyperphosphatemia	Hyperphosphatemia
Hyperphosphatemia	Hyperphosphataemia
Nail Toxicity	Nail bed bleeding
Nail Toxicity	Nail discolouration
Nail Toxicity	Nail disorder
Nail Toxicity	Nail dystrophy
Nail Toxicity	Nail ridging
Nail Toxicity	Nail toxicity
Nail Toxicity	Onychalgia
Nail Toxicity	Onychoclasis
Nail Toxicity	Onycholysis
Nail Toxicity	Paronychia
Nail Toxicity	Onychomadesis
Skin Toxicity	Blister
Skin Toxicity	Dry skin
Skin Toxicity	Erythema
Skin Toxicity	Hyperkeratosis
Skin Toxicity	Palmar erythema
Skin Toxicity	Plantar erythema
Skin Toxicity	Palmar-plantar erythrodysaesthesia syndrome
Skin Toxicity	Rash
Skin Toxicity	Rash erythematous
Skin Toxicity	Rash macular
Skin Toxicity	Rash maculo-papular
Skin Toxicity	Skin atrophy
Skin Toxicity	Skin exfoliation
Skin Toxicity	Skin fissures
Skin Toxicity	Skin lesion
Skin Toxicity	Skin ulcer
Skin Toxicity	Toxic skin eruption
Skin Toxicity	Xeroderma
Dry Mouth	Dry mouth
Dry Mouth	Aptyalism
Mucositis	Mucosal inflammation
Mucositis	Stomatitis

6.7. Appendix 7 FACT-BI Scoring Guidelines (Version 4)

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. Add subscale scores to derive total scores (TOI, FACT-G & FACT-Bl).
5. **The higher the score, the better the QOL.**

Subscale	Item Code	Reverse item?		Item response	Item Score
PHYSICAL WELL-BEING (PWB)	GP1	4	-	_____	=_____
	GP2	4	-	_____	=_____
	GP3	4	-	_____	=_____
	GP4	4	-	_____	=_____
Score range: 0-28	GP5	4	-	_____	=_____
	GP6	4	-	_____	=_____
	GP7	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**PWB subscale score**

SOCIAL/FAMILY WELL-BEING (SWB)	GS1	0	+	_____	=_____
	GS2	0	+	_____	=_____
	GS3	0	+	_____	=_____
	GS4	0	+	_____	=_____
Score range: 0-28	GS5	0	+	_____	=_____
	GS6	0	+	_____	=_____
	GS7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**SWB subscale score**

EMOTIONAL WELL-BEING (EWB)	GE1	4	-	_____	=_____
	GE2	0	+	_____	=_____
	GE3	4	-	_____	=_____
	GE4	4	-	_____	=_____
Score range: 0-24	GE5	4	-	_____	=_____
	GE6	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 6: _____
Divide by number of items answered: _____=**EWB subscale score**

FUNCTIONAL WELL-BEING (FWB)	GF1	0	+	_____	=_____
	GF2	0	+	_____	=_____
	GF3	0	+	_____	=_____
	GF4	0	+	_____	=_____
Score range: 0-28	GF5	0	+	_____	=_____
	GF6	0	+	_____	=_____
	GF7	0	+	_____	=_____

Sum individual item scores: _____

Multiply by 7: _____
Divide by number of items answered: _____=**FWB subscale score**

Subscale	**Item Code**	**Reverse item?**		**Item response**	**Item Score**
BLADDER	BL1	4	-	_____	=_____
CANCER	C2	4	-	_____	=_____
SUBSCALE	C3	0	+	_____	=_____
(BlCS	BL2	4	--	_____	=_____
	C5	4	-	_____	=_____
Score range: 0-48	C6	0	+	_____	=_____
	C7	0	+	_____	=_____
	BL3	4	-	_____	=_____
	BL4	0	+	_____	=_____
	BL5	0	+	_____	=_____
	C8	4	-	_____	=_____
	C9	4	-	_____	=_____

Sum individual item scores:_____
Multiply by 12 : _____
Divide by number of items answered: _____=**BlC Subscale score**

To derive a FACT-Bl Trial Outcome Index (TOI):
Score range: 0-104

_____ + _____ + _____ =_____=**FACT-Bl TOI**
(**PWB** score) (**FWB** score) (**BlCS** score)

To Derive a FACT-G total score:
Score range: 0-108

_____ + _____ + _____ + _____=_____=**FACT-G Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score)

To Derive a FACT-Bl total score:
Score range: 0-156

_____ + _____ + _____ + _____ + _____ =_____=**FACT-Bl Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score) (**BlCS** score)

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www facit.org.

6.8. Appendix 8 OS Assumptions for Sample Size Calculation

Study 42756493BLC3001 is a randomized study of erdafitinib vs standard of care, either chemotherapy (docetaxel or vinflunine in Cohort 1) or pembrolizumab (in Cohort 2), in patients with advanced urothelial cancer and selected FGFR gene aberrations. The primary endpoint is Overall Survival (OS). This appendix describes the rationale supporting the statistical design of the trial, with a focus on median OS estimations.

Median OS for each cohort is derived from the median time from first dose to PD and median time from the PD to death. Time from first dose to death is assumed to be exponentially distributed and the median is approximately calculated as the sum of the time period from first dose to PD and time period from the PD to death. Simulations were used to derive the median OS for each treatment, taking into consideration the proportions of each response category (CR/PR, SD, PD, NE/Unknown) and the median under each category.

Cohort 1: The assumed hazard ratio of 0.65 reflects that the median OS for erdafitinib is estimated to be 9.8 months based on statistical modeling using preliminary data from the phase 2 study, 42756493BLC2001 (FGFR selected population), and the median OS for chemotherapy is estimated to be 6.4 months, derived from modeling using data from the literature (Rosenberg et al., 2016; Bellmunt et al., 2017) and preliminary data from the 42756493BLC2001 study.

Cohort 2: The assumed hazard ratio of 0.69 reflects that the median OS for erdafitinib is estimated to be 10.5 months based on statistical modeling using preliminary data from the phase 2 study, 42756493BLC2001 (selected FGFR population), and the median OS for pembrolizumab is estimated to be 7.24 months, derived from modeling using data from the literature (Rosenberg et al., 2016; Bellmunt et al., 2017) (in an FGFR-unselected population) and preliminary data from the 42756493BLC2001.

7. REFERENCES

Bellmunt J, de Wit R, Vaughn DJ et al. Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. N Engl J Med. 2017 Mar 16;376(11):1015-1026.

Clinical study report for phase 2 study 42756493 BLC2001 "A Phase 2, Two-arm Multicenter, Open-Label Study to Determine the Efficacy and the Safety of Two Different Dose Regimens of a pan-FGFR Tyrosine Kinase Inhibitor JNJ 42756493 in Subjects with Metastatic or Surgically Unresectable Urothelial Cancer with FGFR Genomic Alterations".

Food and Drug Administration (FDA) Type C meeting written responses regarding the proposed Patient Reported Outcome (PRO) endpoints in the study 42756493BLC3001, Application Number 117490, Correspondence Reference ID 4260896, 11 May 2018.

Herdman M, Gudex C, Lloyd A, et al. Development and Preliminary Testing of the New Five-Level Version of EQ-5D (EQ-5D-5L); Qual Life Res; 2011;20(10):1727-1736.

Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics; 1979; **6** (2): 65–70.

Robins JM, Finkelstein DM. Correcting for Non-compliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-rank Tests. Biometrics; 2000; 56:779-788.

Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, phase 2 trial. Lancet. 2016 May 7;387(10031):1909-20.

Tang and Geller. Closed Testing Procedures for Group Sequential Clinical Trials with Multiple Endpoints. Biometrics; 1999, Dec; 55(4):1188-92.