

Integrated Analysis Plan

Study Number: MS100070_0119

Clinical Study Protocol Title: A Phase II, Multicenter, Randomized, Open Label, Parallel-Arm, Umbrella Study of Avelumab (MSB0010718C) in Combination with Other Anti-Tumor Agents as a Maintenance Treatment in Participants with Locally Advanced or Metastatic Urothelial Carcinoma Whose Disease Did Not Progress with First Line Platinum-Containing Chemotherapy

Study Phase: Phase II

Merck Compound: Avelumab

Protocol Version: 25 September 2024/Version 4.0

Integrated Analysis Plan
Author:

Coordinating Author	
Biostatistics, EMD Serono	PPD
Function	Author(s) / Data Analyst(s)
Biostatistics, PPD	PPD
Biostatistics, PPD	PPD
Pharmacokinetics, PPD	PPD
Clinical PK/PD Scientist, Merck Healthcare KGaA	PPD

Statistical Analysis Plan Date and Version: 02 April 2025/Version 3.0

Statistical Analysis Plan
Reviewers:

Function	Name
Senior Biostatistical Reviewer, PPD	PPD
Statistical Lead Programmer, Merck Healthcare KGaA	PPD
Medical Lead, Merck Healthcare KGaA	PPD
Clinical Research Scientist, Merck Healthcare KGaA	PPD
Global Patient Safety Lead, Merck Healthcare KGaA	PPD
New Biological Entities, Drug Metabolism and Pharmacokinetics, EMD Serono	PPD
Clinical PK/PD Scientist, Merck Healthcare KGaA	PPD

Confidential

This document is the property of Merck KGaA, Darmstadt, Germany, or one of its affiliated companies.
It is intended for restricted use only and may not - in full or part - be passed on, reproduced, published or used without
express permission of Merck KGaA, Darmstadt, Germany or its affiliate.

Copyright © 2025 by Merck KGaA, Darmstadt, Germany or its affiliate. All rights reserved.

Approval Page

Integrated Analysis Plan: MS100070_0119

A Phase II, Multicenter, Randomized, Open Label, Parallel-Arm, Umbrella Study of Avelumab (MSB0010718C) in Combination with Other Anti-Tumor Agents as a Maintenance Treatment in Participants with Locally Advanced or Metastatic Urothelial Carcinoma Whose Disease Did Not Progress with First Line Platinum-Containing Chemotherapy

Approval of the Integrated Analysis Plan (IAP) by all Merck Data Analysis Responsible has to be documented within Electronic Document Management System (EDMS) via eSignature. With the approval, the Merck responsible for each of the analysis also takes responsibility that all reviewers' comments are addressed adequately.

By using eSignature, the signature will appear at the end of the document.

1 Table of Contents

1	Table of Contents.....	4
2	List of Abbreviations and Definition of Terms	8
3	Modification History	12
4	Purpose of the Integrated Analysis Plan.....	14
5	Objectives and Endpoints	15
6	Overview of Planned Analyses.....	19
6.1	Analyses for DMC meetings	20
6.2	Interim Analysis.....	20
6.3	Primary Analysis	21
7	Changes to the Planned Analyses in the Clinical Study Protocol	21
7.1	COVID-19 Impact	21
8	Analysis Sets and Subgroups.....	22
8.1	Definition of Analysis Sets	22
8.2	Subgroup Definition and Parameterization	23
9	General Specifications for Data Analyses	25
9.1	Data Handling After Cut-off Date	27
9.2	Definition of Baseline and Change from Baseline	27
9.3	Study Day	27
9.4	Definition of Duration and ‘Time Since’ Variables	28
9.5	Conversion Factors	28
9.6	Date of Last Contact	28
9.7	Definition of On-treatment Period.....	29
9.8	Exposure Time.....	29
9.9	Imputation of Missing Data	30
9.10	Implementation of JB100 Data.....	33
9.10.1	Selection of Participants from JB100	33
9.10.2	Comparability Assessment	33
9.10.3	Propensity Score	33
9.10.4	Preference Score	34

9.10.5	Propensity Score-based Weights	34
9.10.6	Harmonization of Data	35
9.11	Overview of Planned Analyses	35
10	Study Participants	37
10.1	Disposition of Participants and Discontinuations	37
10.2	Important Protocol Deviations	41
11	Demographics and Other Baseline Characteristics	42
11.1	Demographics	42
11.2	Medical History	43
11.3	Other Baseline Characteristics	44
11.3.1	Disease History	44
11.3.2	Nicotine Usage	45
11.4	Prior Anticancer Therapy	45
12	Previous or Concomitant Therapies/Procedures	46
12.1.1	Previous and Concomitant Medications	46
12.1.2	Premedication	46
12.1.3	Previous and Concomitant Procedures	47
12.1.4	Subsequent Anticancer Therapy	47
13	Study Intervention: Compliance and Exposure	48
13.1	Dose Adjustments	49
13.2	Therapy Delays	50
13.3	Exposure	51
14	Efficacy Analyses	53
14.1	Primary Estimand: PFS	53
14.1.3	Supplementary Analyses of the Primary Estimand PFS	59
14.2	Further Estimands	59
CCI	[REDACTED]	
15	Safety Analyses	67
15.1	Adverse Events	67
15.1.1	All Adverse Events	68
15.1.2	Adverse Events Leading to Discontinuation of Study Intervention and Other Actions	70

15.2	Deaths, Other Serious Adverse Events, and Other Significant Adverse Events	72
15.2.1	Deaths	72
15.2.2	Serious Adverse Events	72
15.2.3	Other Significant Adverse Events	73
15.3	Clinical Laboratory Evaluation	77
15.4	Vital Signs	80
15.5	Other Safety or Tolerability Evaluations	81
15.5.1	ECG	81
16	Analyses of Other Endpoints/Estimands	83
16.1	Pharmacokinetics	83
16.1.1	Descriptive Statistics of PK Concentration Data	83
16.1.2	Descriptive Statistics of PK Parameter Data	83
16.1.3	General Specifications for PK Concentration and PK Parameter Data	84
16.1.4	Estimation of Pharmacokinetic Parameters	85
16.1.5	Estimation of Pharmacokinetic Parameters in Serum or Plasma	86
16.1.6	Presentation of PK Concentration and PK Parameter Data	89
16.1.6.1	Listings and Tables	89
16.1.6.2	Graphical Summaries and Individual Plots	90
CCI		
16.2.1	Antidrug Antibody	90
16.2.2	Neutralizing Antidrug Antibody	92
CCI		
17	References	95
18	Appendices	96
Appendix 1 – SAP for DMC		96
Appendix 2 – JBM vs JB100 Eligibility Criteria		97
Appendix 3 – NCCN FACT FBLSI-18 Scoring		106

Appendix 4 – NCI-CTC Gradable and Non-Gradable Safety Laboratory Test Parameters and Direction(s) of Abnormality.....	108
--	-----

2 List of Abbreviations and Definition of Terms

AE	Adverse Event
AESI	Adverse Event of Special Interest
ADA	Anti-Drug Antibody
ASTCT	American Society for Transplantation and Cellular Therapy
AUC	Area Under the Curve
AUC _{0-<i>t</i>last}	AUC from time zero (= dosing time) to the time of the last quantifiable concentration (<i>t</i> _{last})
AUC _{0-∞}	AUC from time zero (= dosing time) up to infinity with extrapolation of the terminal phase
AUC _τ	AUC over the dosing interval τ
BLQ	Below Lower Limit of Quantification
BOR	Best Overall Response
CDISC	Clinical Data Interchange Standards Consortium
C _{eo}	Observed Concentration at the end of the infusion period
CFB	Change from Baseline
CI	Confidence Interval
CL	Total Body Clearance following intravenous administration
(e)CRF	(electronic) Case Report Form
CRS	Cytokine Release Syndrome
CSR	Clinical Study Report
C _{trough}	Concentration observed at the end of a dosing interval immediately before next dosing
CV%	Coefficient of Variation
CCI	
DC	Disease Control
DoR	Duration of Response
DRS-P	Disease Related Symptom-Physical Subscale
EEA	European Economic Area
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status

eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EDMS	Electronic Document Management System
EOI	End of Infusion
EOT	End of Treatment
EQ-5D-5L	EuroQol 5 Dimension 5 Level Scale
FAS	Full Analysis Set
GeoCV%	Geometric Coefficient of Variation
GeoMean	Geometric Mean
HRQOL	Health-related Quality of Life
ICH	International Conference on Harmonization
DMC	Data Monitoring Committee
HR	Hazard Ratio
IA	Interim Analysis
irAE	Immune-related Adverse Event
IRR	Infusion-related Reaction
IWRS	Interactive Web Response System
JB100	JAVELIN Bladder 100
JBM	JAVELIN Bladder Medley
LS	Least Squares
λ_z	Terminal First Order (Elimination) Rate Constant
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum
MMRM	Mixed-effect Model for Repeated Measures
nAB	Neutralizing Antibody
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events
NCCN FACT FBISI-18	National Comprehensive Cancer Network functional assessment of cancer therapy bladder symptom index
ne	Not Evaluable
NED	No Evidence of Disease

CCI	
OR	Objective Response
OS	Overall Survival
PA	Primary Analysis
PD	Progressive Disease CCI
PFS	Progression Free Survival
PRO	Patient-reported Outcome
PT	Preferred Term
PK	Pharmacokinetics
PKAS	PK Analysis Set
Q1W	Every Week
Q2W	Every Two Weeks
Q4W	Every Three Weeks
QoL	Quality of Life
RECIST	Response Evaluation Criteria in Solid Tumor
RMST	Restricted Mean Survival Time
SAP	Statistical Analysis Plan
SAF	Safety Analysis Set
SAE	Serious Adverse Event
SCR	Screening Analysis Set
sCSR	Supplemental Clinical Study Report
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SG	Sacituzumab Govitecan
SMQ	Standardized MedDRA Query
SOC	System Organ Class
t _{1/2}	Terminal Half-life
TEAE	Treatment-emergent Adverse Event
TLF	Tables, Listings, and Figures
UGT 1A1	UDP Glucuronosyltransferase 1A1
ULN	Upper Limit of Normal

VAS	Visual Analogue Scale
V_{ss}	Apparent Volume of Distribution at steady state following intravenous administration
V_z	Volume of Distribution during the terminal phase following intravenous administration

3 Modification History

Unique Identifier for Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	19 June 2023	PPD [REDACTED] [REDACTED] [REDACTED] [REDACTED]	NA – First version
2.0	31 May 2024	PPD [REDACTED] [REDACTED]	- Section 5: DOR endpoint – added in the population definition that only subjects with confirmed OR are included. - Section 7: Updated to indicate that only one interim and one primary analysis will be performed. - Section 7.1: Added COVID-19 vaccines - Section 8.2: Updated the derivation of some subgroups according to what is available in the data - Section 9.2: Specified that JBM and JB100 do not always follow the same rules in terms of baseline definition. - Section 9.3: Added definition for study intervention and analysis relative day. - Section 9.5: Added how to analyze cigarillos. - Section 9.9: Specified the start of subsequent anticancer treatment for JB100 - Section 9.10.2: Removed reference to specific number of subjects randomized overall and in Group A - Section 9.10.3: Specified that the regression model is a logistic mode and added sentence to describe how missing values will be handled. - Section 9.10.5: Specified the meaning of <i>treat</i> - Section 9.11: Specified which data from JB100 will be used for the analyses. - Section 9.12: Removed the section as time windowing is no longer needed. - Section 10.1: Updated the analysis for the compliance with the completion of HRQoL questionnaires. - Section 11.1: Specified how to present race for JB100. Added European Economic Area. Added creatinine clearance to the demographic listing. - Section 11.3.1: Updated definition of CCI [REDACTED] - Section 11.4: Rules provided for the derivation of duration of prior therapy and time from last dose of prior therapy to first dose of study treatment. - Section 12.1.1: Removed sentence referring to multiple ATCs classes. Agreed to present the primary one. - Section 12.1.4: Updated the lists of immune check-point inhibitors and subsequent anticancer drugs.

Unique Identifier for Version	Date of SAP Version	Author	Changes from the Previous Version
			- Section 13.3: Added definition of cycle. Clarified derivation cumulative dose, dose intensity and relative dose intensity for avelumab in JB100. - Section 14: Added that only investigator tumor assessments will be considered. Added that PFS and OS analyses will be repeated without JB100 data in case JB100 is considered suitable. - Section 14.1.1: Added details for the LOESS curve. - Section 14.1.2: Added Sensitivity Analysis 5. - Section 14.2.2: Added derivation of BOR for subjects with no tumor lesion at baseline. - Section 14.2.3: Added swimmer plot. - Section 14.2.4: Added Sensitivity Analysis 1 and 2 and explained how to handle the EOT visit. - Sections 9, 15.1.1 and 15.2.2 updated to add TEAE tables for DMC open session. - Section 15.2.3.4: Added more details for the analysis of cytokine release syndrome events. - Section 15.3: Added INR parameter as gradable parameter. Specified which visits are going to be presented from JB100 data in the boxplots. - Section 15.5.1: Added more details on what to present in table and listing. - Section 16.1: Clarification of PK handling of protocol deviations and correction of inconsistencies. - Appendix 2: Updated comment on Criterion 3. - Appendix 5: Removed as not needed.
3.0	02 April 2025	PPD	- Section 5: Corrected mistake in population definition where additional text was added to OR instead of DoR endpoint. - Section 6.2: Text updated to align with V4 of study protocol where unblinding is planned after the DMC meeting evaluating the results of the interim analysis. - Section 7: Removed text indicating that only one interim and one primary analysis will be performed since updated in V4 of study protocol. - Section 9.2: Added specification for assessments to be considered for the baseline derivation. - Section 9.11: For previous and concomitant medications and procedures, replaced FAS with SAF. Clarified the use of PK data from JB100 study within the immunogenicity analysis. - Section 10.1: Added details for the bar chart for Week 1 Day 1 and EOT visits. - Section 11.4: Updated derivation of duration of prior anticancer drug therapy to align with JB100 study. - Section 12.1.1, 12.1.2 and 12.1.4: Removed the presentation of ATC classification in tables.

Unique Identifier for Version	Date of SAP Version	Author	Changes from the Previous Version
			• Section 15.2.3.1: For interim analysis only, added listing of IRRs with details on premedication and timing of infusion. • Section 14.2.4 and 14.2.5: Updates made to clarify the analysis. • Section 15.5.1: updated to add QT parameter to descriptive statistics as well as Week 1 Day 1 Predose visit. • Section 16.1.5: Text updated for improved clarity and analysis updated to adapt to the available data. • Section 16.2.5: Specified that for SG PK concentration, the analyte hRS7 IgG and hRS7-SN38 will be considered.

4 Purpose of the Integrated Analysis Plan

The purpose of this IAP is to document technical and detailed specifications for the interim and primary analyses of data collected for protocol MS100070_0119. Results of the interim analysis described in this IAP will be provided to the internal Data Monitoring Committee (DMC), results of the primary analysis will be included in the Clinical Study Report (CSR). Additionally, the planned analyses identified in this IAP may be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective IAP will be clearly identified in the CSR.

Additional analyses not covered by this IAP like, for instance, CCI analyses not pre-defined in the protocol or additional subgroup analyses planned for Health Technology Assessment submission will be listed in the respective Data Analysis Responsibility form which is stored in the EDMS together with this IAP.

The IAP is based upon Section 9 (Statistical Considerations) of the study protocol and protocol amendments and of the DMC Charter and is prepared in compliance with ICH E9. It describes analyses planned in the protocol and protocol amendments, with the exception of additional CCI and CCI analyses, which will be described and reported separately. Details of the DMC analyses for quarterly review of the participants' safety without formal statistical analysis are provided in [Appendix 1](#).

The wording used in this IAP is chosen to best match the respective wording in the study protocol template, the CSR template, Clinical Data Interchange Standards Consortium (CDISC) requirements and special requirements for table layouts. Therefore, the following approach is used:

Generally, the term 'participant' will be used instead of 'subject' or 'patient'. However, in tables and listings the term 'subject' will be used to match CDISC requirements, except for in-text tables where 'participant' will be used to match the CSR and protocol templates. Similarly, the term

‘study intervention’ will be used in this document instead of ‘treatment’ to match protocol and CSR templates, however, tables and listings may also use ‘treatment’ for brevity reasons. Exceptions from this rule are commonly used terms like “on-treatment”, “treatment-emergent”, “treatment policy”, “subject-years”, “by-subject”, or names of eCRF pages like “Treatment Termination” page.

5 Objectives and Endpoints

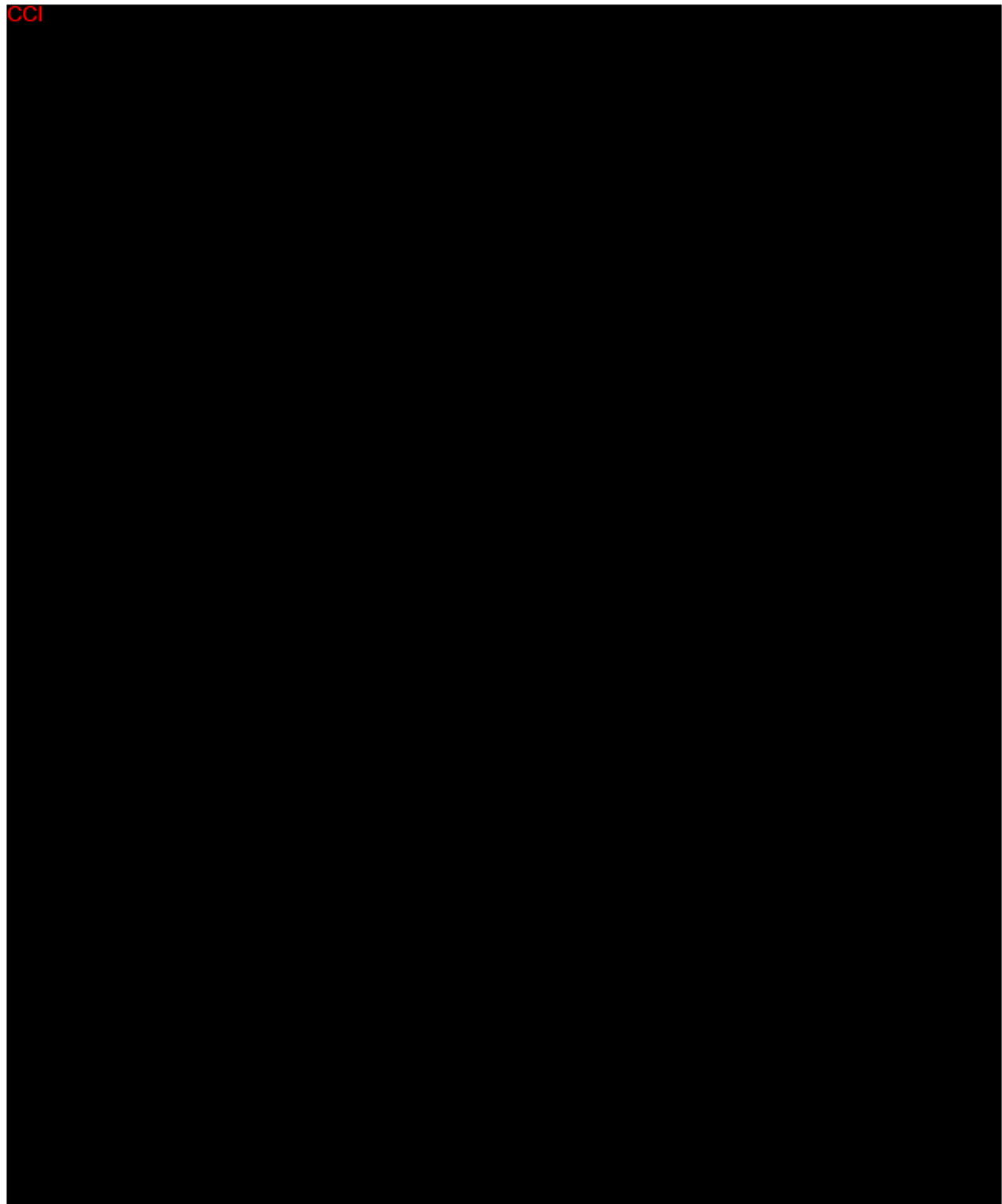
Objectives	Endpoints	Further Estimand Attributes
Primary		
To evaluate PFS of maintenance treatment with avelumab treatment combination regimens compared to maintenance avelumab monotherapy.	PFS as defined from date of randomization to PD according to RECIST v1.1 as assessed by Investigator or death, occurring within 2 scheduled tumor assessments after last evaluable assessment or randomization.	Population: Participants with locally advanced or metastatic urothelial carcinoma whose disease did not progress with first line platinum-containing chemotherapy. Treatment: Maintenance treatment with avelumab treatment combination regimens compared to maintenance avelumab monotherapy. Strategy for handling intercurrent events: - Death within 2 scheduled tumor assessments after last evaluable assessment or randomization will be considered as event (composite strategy). - Discontinuation of treatment: The endpoint will be analyzed regardless of whether or not treatment has been discontinued (treatment-policy strategy). - Start of subsequent anticancer treatment: Tumor assessments after start of subsequent anticancer treatment will not be considered for analysis (hypothetical strategy). Population level summary: Hazard ratio See IAP Section 14.1.1.
To evaluate the safety and tolerability of avelumab combination regimens.	Occurrence of TEAEs, treatment-related AEs, and AESIs as per Qualitative Toxicity Scale (NCI-CTCAE 5.0).	See IAP Section 15.1.
Secondary		
To evaluate OS of maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy followed by subsequent anticancer treatment.	OS as measured by time from randomization to death.	Population: Participants with locally advanced or metastatic urothelial carcinoma whose disease did not progress with first line platinum-containing chemotherapy. Treatment: Maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy.

Objectives	Endpoints	Further Estimand Attributes
		Strategy for handling intercurrent events: The endpoint will be analyzed regardless of whether or not the following intercurrent events had occurred (treatment-policy strategy): • Treatment discontinuation. • Start of subsequent anticancer therapy. Population level summary: Hazard ratio. See IAP Section 14.2.1.
To evaluate OR of maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy.	OR according to RECIST v1.1 assessed by Investigator.	Population: Participants with locally advanced or metastatic urothelial carcinoma whose disease did not progress with first line platinum-containing chemotherapy. Treatment: Maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy. Strategy for handling intercurrent events: • Discontinuation of treatment: The endpoint will be analyzed regardless of whether or not treatment has been discontinued (treatment-policy strategy). • Start of subsequent anticancer treatment: Tumor assessments after start of subsequent anticancer treatment will not be considered for analysis (while not treated with subsequent anticancer therapy strategy). • Progression according to RECIST v1.1 (while not progressed strategy). Population level summary: Objective response rate, odds ratio. See IAP Section 14.2.2.
To evaluate DoR of maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy.	DoR according to RECIST v1.1 assessed by Investigator defined as time from first documentation of objective response to PD or death, occurring within 2 scheduled tumor assessments after last evaluable assessment or randomization.	Population: Participants with locally advanced or metastatic urothelial carcinoma whose disease did not progress with first line platinum-containing chemotherapy and confirmed OR according to RECIST 1.1 as assessed by Investigator. Treatment: Maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy. Strategy for handling intercurrent events: Same strategies as defined for PFS. Population level summary: Median DoR.

Objectives	Endpoints	Further Estimand Attributes
		See IAP Section 14.2.3.
To collect PK concentration data of avelumab, M6223, SG (derived from total SN-38 and unconjugated SN-38), and NKTR-255 to contribute to respective PopPK analyses and exposure-response analyses.	Concentration will be listed and descriptively summarized as appropriate. Population PK and exposure-response results will be reported separately in a standalone report, as it may also include data from other studies.	See IAP Section 16.1.
To characterize the immunogenicity of avelumab administered alone or in combination with SG, NKTR-255 or M6223. To characterize the immunogenicity of M6223, SG or NKTR-255 in combination with avelumab.	Immunogenicity of avelumab, SG, NKTR-255 and M6223, as measured by ADA assays.	See IAP Section 16.2.
To evaluate the effect of maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy on disease related physical symptoms during treatment.	Change from baseline score of NCCN FACT FBISI-18 DRS-P at Week 13 visit.	Population: Participants with locally advanced or metastatic urothelial carcinoma whose disease did not progress with first line platinum-containing chemotherapy. Treatment: Maintenance treatment with avelumab combination regimens compared to maintenance avelumab monotherapy. Strategy for handling intercurrent events: - Discontinuation of treatment: The endpoint will be analyzed according to the hypothetical strategy, i.e., as if participants were still on randomized treatment at Week 13 visit. - Death: The endpoint will be analyzed according to the hypothetical strategy, i.e., as if participants were still alive and on randomized treatment at Week 13 visit. Population level summary: Difference in least-squares means of the change from baseline DRS-P score at Week 13 visit based on a MMRM. See IAP Section 14.2.4

CCI

CCI



Objectives	Endpoints	Further Estimand Attributes
ADA = antidrug antibody; AE = adverse event; AESI = adverse event of special interest; CCI DoR = duration of response; DRS-P = disease related symptom-physical subscale; CCI MMRM = mixed effects model for repeated measures; NCCN FACT FBIS1-18 = National Comprehensive Cancer Network functional assessment of cancer therapy bladder symptom index; NCI-CTCAE = Northern Cancer Institute-Common Terminology Criteria for Adverse Events; CCI; OR = objective response; OS = overall survival; PD = progressive disease, CCI PFS = progression-free survival; PK = pharmacokinetics; QoL = quality of life; RECIST = response evaluation criteria in solid tumors; SG = sacituzumab govitecan; TEAE = treatment-emergent adverse event; CCI CCI		

C
C
I
[Redacted text block]

- I [Redacted]
- I [Redacted]
- I [Redacted]
- I [Redacted]

[Redacted text block]

[Redacted text block]

20/109

11

Changes to the Planned Analyses in the Clinical Study Protocol

COVID-19 Impact

No changes to the planned analysis of the efficacy or safety endpoints are planned due to the impact of Coronavirus disease 2019 (COVID-19) outbreak.

Sensitivity analyses may be considered depending on the observed COVID-19 findings.

A listing will be generated with a description of the impact by COVID-19 on the study for the participants in the FAS. The following findings due to COVID-19 will be presented:

- Adverse events
- Protocol deviations (important and non-important)
- Missed visits
- Missed tumor evaluations
- Any COVID-19 vaccine received (before or during the study)
- Tele-visits (remote audio video or telephone call) or emails replacing site visits
- Drug administration - missed doses
- Drug administration - dose adjustments
- Treatment discontinuation of avelumab and/or combination drug
- Study discontinuation
- Death

8 Analysis Sets and Subgroups

8.1 Definition of Analysis Sets

Screening Analysis Set (SCR)

The Screening Analysis Set (SCR) will include all participants who signed the informed consent, regardless of the participant's randomization and study intervention status in the study.

Full Analysis Set (FAS)

The Full Analysis Set (FAS) will include all randomized participants. Participant will be analyzed according to the study intervention assigned at randomization as per the intent-to-treat principle.

In case the JB100 data is suitable to extend the control group, the FAS will additionally include all participants from the external data set fulfilling the in-/exclusion criteria defined for this protocol and who were randomized to the investigational arm of the JB100 study (i.e., avelumab plus best supportive care).

Safety Analysis Set (SAF)

The Safety Analysis Set (SAF) will include all participants who were administered any dose of any study intervention. Participants will be analyzed per the actual study intervention they received.

In case the JB100 data is suitable to extend the control group, the SAF will additionally include all participants from the external data set fulfilling the in-/exclusion criteria defined for this protocol and who received at least one dose of avelumab in the investigational arm of the JB100 study (i.e., avelumab plus best supportive care).

PK Analysis Set (PKAS)

The PK Analysis Set (PKAS) is a subset of the SAF and will consist of all participants, who receive at least one dose of study intervention, and provide at least one measurable post-dose concentration for an analyte. A measurement below lower limit of quantification (BLQ) is considered a valid measurement. Participants will be analyzed per the actual study intervention they received.

Immunogenicity Analysis Set (IMM)

The Immunogenicity Analysis Set (IMM) will include all participants who receive at least one dose of study intervention and have at least one valid ADA result. Participants will be analyzed per the actual study intervention they received.

In case the JB100 data is suitable to extend the control group, the IMM will additionally include all participants from the external data set fulfilling the in-/exclusion criteria defined for this protocol, who received at least one dose of avelumab in the investigational arm of the JB100 study (i.e., avelumab plus best supportive care), and who have at least one valid ADA result.

Analyses per Analysis Set

See Table 3 Overview of Planned Analyses: Analysis Set, Display, and the Usage of JB100 Data in Section 9.11.

8.2 Subgroup Definition and Parameterization

Subgroup analyses will be performed on primary and certain secondary efficacy endpoints using the subgroups as defined in Table 1. Endpoints for which subgroup analyses will be performed can be found in Section 8.2. All subgroup analyses will be **CCI**, no adjustment for multiplicity will be performed.

For the definition of subgroup level, data as documented in the electronic Case Report Form (eCRF) or provided by Vendors and made available in the SDTM datasets will be taken. The category "missing" will not be included in any subgroup analysis.

In case of low number of participants within a category (< 5 participants in any of the arms), categories will be pooled when meaningful.

Table 1. Subgroup Definition (with Reference Level)

Subgroup	Subgroup level	Definition/ Derivation
Age	< 65 years (reference level) ≥ 65 years	Collected on the "Demographics" eCRF page. For JB100, the variable AAGE in ADaM.ADSL will be used.
Sex	- Male (reference level) - Female	Collected on the "Demographics" eCRF page. "Undifferentiated" will be handled as "missing" (only for the purpose of the subgroup analysis). For JB100, sex is available in SDTM.DM.
Race	- White (reference level) - Black or African American - Asian - Other/More than one race	Collected on the "Demographics" eCRF page. If race is "Not permitted per local regulation" then it will be handled as "missing" (only for the purpose of the subgroup analysis). "Chinese", "Japanese", "Korean" and "Other Asian" will all be presented under "Asian". In case, for example, the race indicated is both "White" and "Japanese", the participant will be included in the "Other/More than one race" category. For JB100, the variable ARACE in ADaM.ADSL will be used.
Pooled geographic region	- North America - Europe (reference level) - Asia - Rest of the World	This is derived considering the country where each site is located. The variable "COUNTRY" will be provided in the SDTM demographic dataset.

Subgroup	Subgroup level	Definition/ Derivation
Visceral metastases at baseline	- Present - Absent (reference level)	As indicated at the time of the randomization in the IWRS system. For JB100, this was also a stratification factor for the study and the variable STRAT12 in ADaM.ADSL will be used.
CCI status at baseline	- Positive (reference level) - Negative	The CCI testing will be performed by immunohistochemistry. See Section 11.3.1 for more details on the derivation. For JB100, the variable CCI in ADaM.ADSL will be used. The "Unknown" value will be treated as if the value is missing.
First-line chemotherapy regimen	- Gemcitabine + Cisplatin (reference level) - Gemcitabine + Carboplatin	Refer to the "Prior anti-cancer drug therapies details" eCRF page where the intent of therapy is "Metastatic/Locally advanced" and the line number is 1. For JB100, the variable FLCTREG in ADaM.ADSL will be used. NOTE: in case of participants switching from Cisplatin to Carboplatin or vice versa, the first drug received will be considered to assign the participants to one of those levels.
Best response to first-line chemotherapy	- CR or PR - SD (reference level)	Refer to the "Prior anti-cancer drug therapies details" eCRF page: it is the best response to the first-line chemotherapy regimen. For JB100, this was a stratification factor for the study and the variable STRAT11 in ADaM.ADSL will be used.
ECOG PS	0 (reference level) ≥ 1	Refer to "ECOG Performance Status" eCRF page. For JB100, ECOG assessments are available in SDTM.QSEO.
Creatinine clearance at baseline	≥ 60 ml/min (reference level) < 60 ml/min	Baseline value to be derived using the formula specified in Section 15.3. For JB100, creatinine values are available in SDTM.LB.
Liver lesions at baseline	- No (reference level) - Yes	Refer to "Disease History - Urothelial Carcinoma" eCRF page. For JB100, the variable LVRLESBL in ADaM.ADSL will be used.
Lung lesions at baseline	- No (reference level) - Yes	Refer to "Disease History - Urothelial Carcinoma" eCRF page. For JB100, the variable LNGLESBL in ADaM.ADSL will be used.

9 General Specifications for Data Analyses

The results of this study will be reported using summary tables, figures, and data listings, as appropriate. All data will be summarized by study intervention and/or scheduled time point, as applicable.

TLFs for DMC at Interim Analysis

Open session tables for the interim analysis will be presented with intervention groups combined into an overall column. Closed session TLFs will present the data for each intervention group and if applicable, combined. No masking will be applied.

For the open session, the following tables will be prepared:

- Subject disposition
- Important protocol deviations
- Disease history
- Treatment emergent adverse events (see Section 15.1.1 and 15.2.2 for more details)

Study Intervention Group Names for TLFs

In TLFs, the study intervention groups will be referenced as follows:

- Group A: avelumab
- Group B: avelumab + SG
- Group C: avelumab + M6223
- Group D: avelumab + NKTR-255

Group B, C and D will be referenced to as the combination groups. Drugs SG, M6223 and NKTR-255 will be referenced to as the combination drugs.

In case the external JB100 data will be considered suitable for the analyses, data for these participants will be presented in the following two ways:

- in outputs where the external control group is presented separately, the external control group will be referenced as “JB100”
- in outputs where the external control group is added to Group A, the extended control group will be referenced to as “Group A + JB100”

Listings

In the individual participant data listing all individual data collected or derived will be listed. Unscheduled measurements will be included in the listings. All listings will be sorted by

intervention group (i.e., Group B, C, D, A and, if applicable, JAVELIN 100), subject ID, and/or nominal time point, as appropriate. The external JB100 data will be presented after Group A data.

Tables and Descriptive Statistics

All data will be summarized by intervention group and nominal time point, as appropriate. The overall total column combining the data from all the intervention groups may be also presented, if applicable. In general, the order of the Groups should be Group B, Group C, Group D, Group A, JB100, or Group A + JB100, as applicable.

Unless otherwise specified, continuous variables will be summarized using descriptive statistics, i.e., the number of participants with non-missing values (n), mean, standard deviation, median, 25th percentile (Q1) and 75th percentile (Q3), minimum, and maximum.

Qualitative variables will be summarized by frequency counts and percentages. Unless otherwise stated, the calculation of proportions will be based on the number of participants of the analysis set of interest. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

For variables where a participant may have more than one category due to multiple responses per participant, the number of participants included in each category will be summarized as a percentage from all participants. Therefore, the total frequency across categories may not equal the total number of all participants in that analysis group.

All statistical tests mentioned in this IAP **CCI**. All statistical tests comparing study intervention groups will be performed two-sided. If confidence intervals (CIs) are calculated, these will be two-sided with a confidence probability of 95%.

Scheduled and Unscheduled Visits

Data collected at both scheduled and unscheduled visits will be included in the derivation of safety endpoints.

Descriptive statistics by nominal visit or time point, e.g., for laboratory measurements, will include only data from scheduled visits. Unscheduled visits will be included in the derivation of baseline and worst on-treatment value.

Pooling of Sites

In order to provide overall estimates of study intervention effects, data will be pooled across sites.

Statistical Software

Phoenix® WinNonlin® Version 8.3, or higher (Certara, L.P., Princeton, New Jersey, USA) will be used to derive PK parameters. All other analyses will be performed using SAS® Software Version 9.4 or higher or R (www.r-project.org), Version 3.2.5 or higher.

9.1 Data Handling After Cut-off Date

By its nature, data after cut-off may be incomplete and subject to further change and will not be used for summary statistics, statistical analyses, listings or imputations.

Stop dates are not affected by this rule, e.g., a stop date of Adverse Events (AEs), which starts prior to the cut-off, but stops after date of cut-off, will not be changed.

9.2 Definition of Baseline and Change from Baseline

In general, the last non-missing measurement prior to randomization will serve as the baseline measurement for efficacy (including PRO), safety and immunogenicity analyses. Those values should be also used for subgroup classification. If no such value is available, the last measurement prior to the first study intervention will be used as the baseline measurement with the exception of pre-randomization assessments used for the derivation of efficacy endpoints.

If an assessment that is planned to be performed before randomization, or study intervention per protocol is performed on the same day as the randomization or start of study intervention, respectively, but the time of assessment and/or the time of first dose of study intervention is not available, it will be assumed that it was performed prior and will be considered for the derivation of baseline (except for PRO).

If a scheduled pre-dose measurement occurred post-dose, the corresponding measurement will be analyzed similar to an unscheduled post-dose measurement.

Absolute and percent changes from baseline are defined as

- absolute change = visit value – baseline value
- percent change = $100 * (\text{visit value} - \text{baseline value}) / \text{baseline value}$.

Baseline values should be used for subgroup classification.

Note: JB100 may have used different definitions for the selection of the baseline value (e.g., for safety assessments). If JB100 data are suitable, the same rules applied for JBM data will be followed.

9.3 Study Day

Study Day 1 is the day of randomization, the day before is Day -1 (no Day 0 is defined).

Study intervention Day 1 is the day of start of study intervention, the day before is Day -1 (no Day 0 is defined).

The analysis relative day will be calculated using the date of randomization as Day 1 when presenting results related to subject disposition, protocol deviations and tumor assessments; in all the other type of data, Day 1 will be the start of study intervention.

9.4 Definition of Duration and ‘Time Since’ Variables

Durations in days will be calculated by the difference of start and stop date + 1 (e.g., survival time (days) = date of death – date of randomization + 1) if not otherwise specified.

The time since an event (e.g., time since first diagnosis) will be calculated as reference date minus date of event.

9.5 Conversion Factors

The following conversion factors will be used to convert days into weeks, months or years:

1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

For the calculation of smoking exposure for cigarettes, cigarillos (for JB100 only), cigars and pipes:

- Cigarette equivalents are calculated as follows: 1 cigarillo is regarded equivalent to 1 cigarette, 1 cigar is regarded equivalent to 5 cigarettes and 1 pipe is regarded equivalent to 3 cigarettes.
- Duration of smoking (years): (end date of smoking – start date of smoking + 1) / 365.25. For the imputation of the start and end dates needed for the calculation of the smoking duration refer to Section 9.9.
- Pack-years:
 - calculate cigarette equivalents per day using the conversion factors given above
 - convert to packs per day where 20 cigarettes are regarded as 1 pack
 - pack-years = packs per day × duration of smoking (years)

9.6 Date of Last Contact

The date of last contact will be derived for participants not known to have died at the analysis cut-off using the latest complete date prior to or at the data cut-off date among the following:

- Last known to be alive date collected on the “Subject Status/Survival Follow-up” eCRF page (consider only the dates provided for participants who are known to be alive)
- AE start and end dates
- Date of last study intervention
- Date of last tumor assessment
- Concomitant medications/procedures start and end dates
- Premedication start date
- Subsequent anticancer therapy start and end dates

- Date of discontinuation from the “Study Termination” eCRF page (do not use if reason for discontinuation is lost to follow-up or death)
- All participant assessment dates (blood draws (laboratory, PK, CCI), vital signs, ECOG performance status, electrocardiograms (ECGs), tumor assessments, quality of life questionnaires, pregnancy tests)

Only dates associated with actual examinations of the participant will be used in the derivation (i.e., if the assessment is reported as “Not Done” then the corresponding date should not be used). Dates associated with a technical operation unrelated to participant status such as the date a blood sample was processed will not be used. Assessment dates after the cut-off date will not be considered for the derivation of the last contact date. This same rule is to be applied to any of the start and stop dates mentioned in the list above.

9.7 Definition of On-treatment Period

The on-treatment period is defined as the time from the first dose of study intervention day to the last administration day of study intervention + 30 days, or the cut-off date or death or the start day of subsequent anticancer therapy – 1 day, whichever occurs first.

Subsequent anticancer therapies may be drug therapies, radiotherapies or surgeries recorded in the “Anti-cancer Treatment After Discontinuation Details”, “Radiotherapy After Discontinuation Details” and “Surgery After Discontinuation Details” eCRF pages. The earliest date of any therapy will be selected. Refer to Section 9.9 for the imputation of partial dates.

9.8 Exposure Time

The study intervention in the combination groups will be considered as ended when both drugs are discontinued, otherwise the study intervention will be considered as ongoing.

The duration of exposure to study intervention and to each study drug separately will be calculated in months as:

$$(\text{end date of exposure} - \text{date of first dose} + 1) / 30.4375,$$

where the end date of exposure is as follows:

- Avelumab drug: date of last dose of avelumab + 14 – 1
- SG drug:
 - If the last dose of SG was a Day 1 infusion of a cycle: date of last dose of SG + 7 – 1
 - If the last dose of SG was a Day 8 infusion of a cycle: date of last dose of SG + 14 – 1)
- M6223 drug: date of last dose of M6223 + 14 – 1
- NKTR-255 drug: date of last dose of NKTR-255 + 28 – 1

- Avelumab + combination drug: maximum (avelumab end date of exposure, combination drug end date of exposure)

9.9 Imputation of Missing Data

Except for partially missing dates as described below, all data will be evaluated as observed, and no imputation method for missing values will be used.

Partial dates will be presented in listings in the format “____YYYY” or “__MMYYYY”, as applicable. In all participant data listings presenting adverse events, imputed values will be presented, and imputed information will be flagged.

Missing statistics, e.g., when they cannot be calculated, should be presented as “ne”. For example, if $n=1$, the measure of variability (SD) cannot be computed and should be presented as “ne”.

Imputed dates will be used for the calculation of durations or “time since” variables.

Disease History

Incomplete dates for disease history (e.g., initial diagnosis date, date of documented metastatic disease diagnosis) will be imputed as follows:

- If the day is missing, it will be imputed to the 15th day of the month.
- If both day and month are missing and the year is prior to the year of the first study intervention, the month and day will be imputed as July 1st.
- If both day and month are missing and the year is same as the year of the first study intervention, the month and day will be imputed as January 1st.
- If the date is completely missing, no imputation will be performed.

Adverse Events

Incomplete AE-related dates will be imputed as follows:

- In case the onset date is missing completely or missing partially but the onset month and year, or the onset year are equal to the start of study intervention then the onset date will be imputed by the minimum of start of study intervention and AE resolution date (if not missing).
- In all other cases, the missing onset day or missing onset month will be imputed by 1.
- Incomplete stop dates will be imputed by the last day of the month (if day is missing only), if not resulting in a date later than the date of participant’s death. In the latter case, the date of death will be used to impute the incomplete stop date.
- In all other cases, the incomplete stop date will not be imputed.

Special cases: AEs may be split into multiple records if the toxicity grade and/or the seriousness changes (see Section 15.1 for more details). The end date of the previous record and the start date of the new record are expected to be the same. If the month and year are indeed the same, but the

day is missing, it will be imputed by 15. In case the imputed start date of the new record is after the end date of this same record, no imputation will be performed.

Previous and Concomitant Medication

Incomplete prior/concomitant medication start and stop dates will be imputed as presented in Table 2.

Table 2. Imputation rules for missing/incomplete start/end dates of medication

	Start Date	End Date
Day missing only	Day = 1	Day = Last day of month
Month missing	Day = 1 Month = Jan	Day = 31 Month = Dec
Year missing	Date = Missing No imputation	
All	If imputed date > date of death: imputation by date of death	

For the derivation of previous and concomitant medications, the following rules will be applied (imputed dates are to be considered):

Previous Medication:

- Start date $\leq$ Start of study intervention OR
- Start date = Missing

Concomitant Medication:

- End date $\geq$ Start of study intervention AND (Start date $\leq$ End of study intervention OR Start date = Missing) OR
- End date = Missing AND (Start date $\leq$ End of study intervention OR Start date = Missing)

The derivation is based on the following principles

- Imputation leads to max. reasonable duration
- Worst case: if medication is administered the same day as start of study medication, medication is classified as both concomitant and previous

Prior Anticancer Drug Therapy

Incomplete start and end dates of prior anticancer drug therapy will be imputed as follows and will be used for the derivation of the duration of prior anticancer drug therapy and time from last dose of prior anticancer drug therapy to first dose of study intervention:

- If only day is missing, the start date will be imputed as the first day of the month and the end date as the last day of the month. If the end date falls beyond the randomization date then impute the end date as randomization date – 29.

- If the month is missing or the date is completely missing, no imputation will be done.

Subsequent Anticancer Treatment

Incomplete dates for start date of subsequent anticancer treatment (drug therapy, radiation, surgery) will be imputed as follows and will be used for determining censoring dates for efficacy analyses and in the derivation of the end of on-treatment period:

- If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anticancer treatment is before that date. In that case, the incomplete anticancer treatment start date will be imputed as the end date of the anticancer treatment.
- If both day and month are missing, no imputation will be performed.

Incomplete subsequent anticancer treatment stop dates will not be imputed.

For JB100, the date of start of subsequent anticancer therapy was derived following different imputation rules. It will not be re-derived but rather taken as it is from JB100 ADaM ADDATES.

Tumor Assessments

All investigation dates (e.g., X-ray, CT scan) must be completed with day, month and year.

If there are multiple scan dates associated with an evaluation, i.e., radiological assessments occur over a series of days rather than the same day, the choice of date of assessment could impact the date of progression and/or date of response. If there are multiple scan dates associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.

Death Date

For the purpose of survival analyses partially missing death dates will be imputed as follows:

- If only the day is missing, the death date will be imputed to the maximum of the (non-imputed) day after the date of last contact and the 15th day of the month.
- Otherwise, it will not be imputed.

Smoking Duration

Imputation will be used for the calculation of duration of smoking for cigarettes, cigars and pipes (years).

In case of missing month, it will be imputed as follows:

- Missing month of start date will be imputed as January
- Missing month of end date will be imputed as December

In case the start date is missing completely, the start date will not be imputed, and duration will not be calculated.

In the case that the end date is missing completely as the participant is still a smoker, in order to allow the calculation of duration, the end date will be imputed by the screening informed consent date.

As the day of the month is also needed to calculate the duration of smoking, the first day and last day of the month for the start date and end date will be used, respectively. If the imputed start date is later than the date of screening, the date will be replaced by the date of screening. The same rule will be applied to the end date for former smokers.

9.10 Implementation of JB100 Data

9.10.1 Selection of Participants from JB100

Participants from the JB100 study who were randomized to the avelumab plus best supportive care (BSC) arm, i.e., the investigational arm, and comply with the inclusion and exclusion criteria applied in the JBM study will be selected for analysis. For programming purposes, the inclusion and exclusion criteria of JBM study are compared and matched with those from the JB100 study as described in [Appendix 2](#). Only those participants fulfilling all the eligibility criteria will be selected for the comparability assessment described in Section 9.10.2.

9.10.2 Comparability Assessment

A comparability assessment will be conducted to verify whether the control group from the JBM study could be extended by the selected participants of the JB100 study.

Pre-chosen baseline covariates with possible impact on the PFS and/or OS outcome will be used to calculate the preference score, i.e., a prevalence adjusted propensity score defined as the conditional probability of being a participant of the JB100 study in contrast to being a participant of the JBM study given observed baseline covariates (Walker 2013). The baseline covariates used for the calculation of the preference score are age, visceral metastases, **CCI** status, ECOG PS, liver lesions and lung lesions. Same categorization as described in Section 8.2 will be applied.

Comparability will be examined when all baseline characteristics of the participants of JBM are available or at the latest at the time of the interim analysis.

9.10.3 Propensity Score

For each treatment-control comparison (e.g., participants randomized to Group B compared to participants randomized to Group A extended by participants selected from JB100 as described in Section 9.10.1), the propensity score is estimated using a logistic regression model, in which study affiliation (participants from JBM versus JB100) is regressed on observed baseline characteristics.

The estimated propensity score e is the predicted probability of being part of the JBM study derived from the fitted logistic regression model, i.e.,

$$e_i = P(\text{subject } i \text{ is part of JBM} \mid X_i), \quad i = 1, \dots, N_{\text{exc}} + N_{\text{randC}} + N_{\text{comb}},$$

with

- X_i : observed baseline characteristics of subject i
- N_{exC} : number of participants from external group (i.e., JB100)
- N_{randC} : number of participants in randomized control (i.e., Group A)
- N_{comb} : number of participants in combination group (i.e., Group B or C or D).

In the PROC LOGISTIC model, the option MISSING will be added in order to consider missing values as an additional level of the baseline covariates and avoid the exclusion of subjects from the model (Section 6.1.1 of Stuart E. A., 2010).

9.10.4 Preference Score

The preference score is obtained by subtracting the natural logarithm of the JBM participants' prevalence (i.e., $p = \frac{N_{randC} + N_{comb}}{N_{exC} + N_{randC} + N_{comb}}$) divided by the JB100 participants' prevalence (i.e., $1 - p = \frac{N_{exC}}{N_{exC} + N_{randC} + N_{comb}}$) from the logit of the propensity score:

$$\text{logit}(f) = \ln\left(\frac{f}{1-f}\right) = \ln\left(\frac{e}{1-e}\right) - \ln\left(\frac{p}{1-p}\right)$$

with f as preference score, e as the propensity score, and p as the fraction of participants randomized to the JBM study. By taking the anti-logit (expit) of the result, the preference score f will be obtained as follows:

$$f = \frac{e^{\text{logit}(f)}}{1 + e^{\text{logit}(f)}}, \quad f \in (0, 1).$$

The JB100 data will be declared suitable as external control if more than 50% of the participants in the randomized data (i.e., Group A and Group B or C or D) and more than 50% of the selected participants of the JB100 data have a preference score between 0.3 and 0.7. This must be true for each set of treatment-control comparisons. If suitable, the selected participants from the JB100 study will be added to the randomized control (Group A) and serve overall as an extended control group. Otherwise, the external data will not be used for any analyses described in this IAP.

9.10.5 Propensity Score-based Weights

In case the external control is suitable, the propensity scores will be used as weights in order to minimize the impact of confounding factors on the estimation of causal treatment effects (Austin 2013). Each participant from the JBM study will be assigned a weight of 1, while participants from the external control will get a propensity score-based weight with the sum of weights adding up to the number of participants in the randomized control Group A. This will control the impact of

information gained from JB100 and guarantee equality between external and randomized control data. The weights will be added to the respective statistical analysis of PFS and OS.

Based on the propensity score, the weights are defined as

$$w_i = \mathbb{I}(treat_i = 1) + \mathbb{I}(treat_i = 0, \kappa = randomized) \frac{\mathbb{I}(treat_i = 0, \kappa = external) \frac{e_i}{1 - e_i}}{\sum_j \mathbb{I}(treat_j = 0, \kappa = external) \frac{e_j}{1 - e_j}} \times N_{randc},$$

$$i, j = 1, \dots, N_{exC} + N_{randc} + N_{comb}.$$

Note: $\mathbb{I}(criterion\ 1, criterion\ 2, \dots)$ is an indicator function returning 1 if all the criteria listed are met, 0 otherwise. $treat_i = 1$ indicates that the participant was randomized to either Group B, C or D. $treat_i = 0$ and $\kappa = randomized$ indicates that the participant was randomized to Group A, $treat_i = 0$ and $\kappa = external$ indicated that the participant belongs to JB100.

9.10.6 Harmonization of Data

In the JB100 study, the MedDRA version 22.1 was the last used for coding. In the JBM study, the MedDRA version used will be the latest available at the time of analysis. In case the JB100 data will be used, the AE terms provided by the investigator will be up-versioned. In terms of NCI CTCAE version used for the grading of AEs, JB100 used version 4.03 and JBM will use version 5.0. In this case the AE grades will be analyzed as reported in the data.

Medical history and procedure terms will also be recoded using the MedDRA version applied to JBM at the time of analysis. Similarly, drug terms in JB100 will be recoded using the same WhoDrug version applied to JBM.

Toxicity grading for laboratory parameters in JB100 will be derived using the same rules applied to JBM, which are based on NCI CTCAE version 5.0.

9.11 Overview of Planned Analyses

In case the JB100 data will not be considered suitable to expand control Group A, the JB100 data will not be included in any of the analyses described in this IAP. Otherwise, see details in Table 3 and further information in the respective description of the analysis. In case a specific variable or option of a variable was not collected, e.g., undifferentiated was not an option available for sex in the JB100 eCRF, it will be replaced by “na” (i.e., not applicable) in tables or the analysis will not be done, e.g., specific laboratory parameters. In general, listings will include JB100 data when related tables include the data.

Unless stated otherwise, all analyses apply to both IA and PA.

Since the propensity-score based weights (see Section 9.10.5) are based on covariates with supposedly prognostic and probably predictive impact on PFS and OS, weights will be used for respective endpoints but not for other efficacy or safety endpoints.

Table 3 Overview of Planned Analyses: Analysis Set, Display, and the Usage of JB100 Data

	Analysis	Analysis Set	Display ¹		Use of JB100 data, if suitable	
			By study inter- vention	Total ²	Combined and weights ³	As separate column ⁴
Study Participants	Subject Disposition	SCR/SAF/FAS	✓	✓		✓
	HRQoL Compliance	FAS	✓	✓		
	Protocol Deviations	FAS	✓	✓		
	Inclusion and Exclusion Criteria Deviations JB100	Investigational arm of JB100		✓		
Demographic and Other Baseline Characteristics	Demographics	FAS	✓	✓		✓
	Medical History					
	Other Baseline Characteristics					
	Prior Anticancer Therapy					
Previous and Concomitant Medications, Procedures	Previous and Concomitant Medications	SAF	✓	✓		✓
	Premedication					
	Previous and Concomitant Medications					
	Subsequent Anticancer Therapy					
Study Intervention: Compliance and Exposure	Dose Adjustment	SAF	✓			✓
	Therapy Delays					
	Exposure					
Efficacy Analyses	Primary Estimand: PFS	FAS	✓		✓	
	Secondary Estimand: OS	FAS	✓		✓	
	Secondary Estimand: OR	FAS	✓			✓
	Secondary Estimand: DoR	FAS	✓			✓
	Secondary Estimand: NCCN FACT FBISI-18 DRS-P	FAS	✓			
	████████████████████ ████████████████████ ████████	████	CL			
Safety Analyses	AEs	SAF	✓			✓
	Deaths, Other Serious AEs, and Other Significant AEs					
	Clinical Laboratory Evaluations					
	Vital Signs					

CCI

¹ Display refers to columns with descriptive information, e.g., frequencies. Statistical tests or calculated measurements, e.g., odds ratios, are not described here and details of how results are presented can be found within the respective analysis.

² Regardless of the usage of JB100 data, the total number will include (randomized) JBM data only.

³ In case JB100 is suitable, control Group A and JB100 data will be combined and handled as one control group. Weights will be calculated as described in Section 9.10.5 and used for population level summaries and respective estimations.

⁴ In case JB100 is suitable, JB100 data will be displayed in a separate column.

From JB100, two sets of data are available:

- Overall Survival update (OS update) with cutoff date 04 June 2021
- Supplemental Clinical Study Report (sCSR), coinciding with the final analysis, with cutoff date 06 April 2023

If JB100 will be considered suitable, data from the OS update will be used for the analysis. However, immunogenicity and PK data are not included in the OS data and therefore sCSR data will be used instead; immunogenicity and PK records with a date after 04 June 2021 will be ignored in the analysis. JB100 PK data will only be used within the immunogenicity analysis (see Section 16.2.5).

When deriving variables or parameters for participants coming from the JB100 study that consider the cutoff date, the date 04 June 2021 will be used.

Note: The decision to use data from the OS update instead of the sCSR was based on the former's completeness. For example, the sCSR set lacks efficacy data.

10 Study Participants

The subsections in this section include specifications for reporting participant disposition and study intervention/study discontinuations. Additionally, procedures for reporting protocol deviations are provided.

10.1 Disposition of Participants and Discontinuations

The number and percentage of participants in each of the below disposition categories will be presented by group of study intervention and total, where applicable. Percentages will be presented with respect to the number of randomized participants in the respective group.

Participants Disposition Status

- Total number of participants screened (i.e., participants who gave informed consent)

- Number of participants who discontinued from the study prior to randomization overall and grouped by the primary reason (e.g., the failed specific inclusion or exclusion criteria, withdrawal of consent)
- Number and percentage of randomized participants

The end of study intervention status will be summarized by:

- Number and percentage of randomized participants who did not receive any dose of study intervention
- Number and percentage of randomized participants with ongoing study intervention (i.e., at least one study drug ongoing)
- Number and percentage of randomized participants who discontinued study intervention (i.e., all study drugs discontinued)
- Number and percentage of randomized participants who did not receive any dose of avelumab
- Number and percentage of randomized participants with ongoing avelumab
- Number and percentage of randomized participants who discontinued avelumab (overall and by primary reason)
- Number and percentage of randomized participants who did not receive any dose of the combination drug
- Number and percentage of randomized participants with ongoing combination drug
- Number and percentage of randomized participants who discontinued the combination drug (overall and by primary reason)

The end of study status will be summarized by:

- Number and percentage of randomized participants with ongoing study intervention
- Number and percentage of randomized participants off-treatment and in follow-up (overall and by last known to be alive date compared to cutoff date: within 3 months or more than 3 months)
- Number and percentage of randomized participants who discontinued the study after randomization (overall and by primary reason)

The listing of individual disposition will include all participants in the SCR Analysis Set and will report data related to screening, randomization, study intervention, study participation status and analysis set.

Analysis Sets

Analysis: PA

The number and percentage of participants screened and randomized in each analysis set will be provided overall, by region (North America, Europe, Asia, Rest of the World), by country within region and by site.

Enrollment Details

Analysis: PA

The results of the randomization algorithm (according to IWRS) of participants in the FAS will be summarized as follows:

- Number of randomized participants overall, by region [North America, Europe, Asia, Rest of the World], by country within region and by site.
- Number of randomized participants by randomization strata (IWRS)
- Number of randomized participants by randomization strata (eCRF)
- Cross tabulation: stratum by IWRS vs. stratum by eCRF
- Cross tabulation: randomized vs. treated

A listing reporting the randomization strata according to IWRS and investigator will be produced.

Time to Study Intervention Discontinuation

Analysis: LA

A swimmer plot will be produced to present the time to permanent study intervention discontinuation for all the participants in the SAF analysis set who permanently discontinued study intervention. Time to permanent study intervention discontinuation will be calculated as the time from first dose of any study drug until permanent discontinuation of study intervention as indicated in the study drug termination eCRF pages. For each participant and study drug, the primary reason of treatment discontinuation will be presented.

Health-related Quality of Life (HRQoL) Compliance Over Time

Analysis: PA

Disposition of HRQoL questionnaires will be described. Separate analyses will be provided for CCI

Population-level summary:

- NCCN FACT FBISI-18:

$$\% \text{ Compliance} = 100 \times \frac{\text{number of participants who returned questionnaire}}{\text{number of participants for whom a questionnaire is expected}}$$

- EQ VAS:

$$\% \text{ Compliance} = 100 \times \frac{\text{number of participants who returned questionnaire}}{\text{number of participants for whom a questionnaire is expected}}$$

Compliance rates will be displayed for each instrument by study intervention group and overall, and by baseline assessment and subsequent scheduled visits (data will be presented as collected, no time windowing will be applied).

A questionnaire is expected to be completed at a scheduled assessment time when a participant is available for the planned visit at the scheduled assessment time. Participants are expected to complete questionnaires if they stay on at least one study drug. Visits are not expected after data cut-off (i.e., the participant is still on treatment at the time of analysis but has not reached a specific visit yet), after end of treatment of all study drugs, or every other visit when the schedule changes from Q2W to Q4W for those participants who continue with study treatment after PD. In such situations data cannot be considered as missing.

For non-compliance following summary will be provided by study intervention group and overall, and by baseline assessment and subsequent visits:

- Number and percentage of randomized participants with expected visits:
 - Number and percentage of participants who completed the questionnaire, by timing of completion compared to the drug administration: before start of treatment, after start of treatment, timing of start of treatment missing or treatment not done
 - Number and percentage of participants who did not complete the questionnaire, by reasons for non-completion: visit not done or reason as reported in the “Quality of Life Questionnaire(s)” page in the eCRF
- Number and percentage of randomized participants with non-expected visits:
 - Baseline visit: reasons for visit not expected due to study discontinuation before Week 1 Day 1 visit (i.e., lost to follow-up, protocol non-compliance, death, withdrew consent)
 - Any post-baseline visit (end of treatment (EOT) visit excluded):
 - Questionnaire completed but not due (i.e., subjects with progressive disease who kept completing the questionnaires every two weeks instead of four)
 - Questionnaires not expected due to data cut-off, treatment discontinuation (overall and by all reasons given in the eCRF: due to AE, lost to follow-up, protocol non-compliance, death, progressive disease, withdrawal of consent, and other), and other (e.g., due to change in assessment schedule after progressive disease, due to study discontinuation without having been treated)
 - EOT visit: never started treatment, treatment ongoing, treatment discontinuation due to lost to follow-up, death, withdrawal of consent.

A 100% stacked bar chart for participant disposition regarding patient-reported outcome (PRO) questionnaires' expectations will be presented summarizing the percentage of participants being expected to complete the questionnaires at the respective visit and the percentage of participants being not expected to complete any questionnaire. The percentage of participants with non-expected visits will be presented by reason: data-cut, death, treatment discontinuation due to AE, progressive disease, or other reasons (including withdrawal, lost to follow-up, protocol non-compliance, other), and other. At Week 1 Day 1, questionnaires are always expected except for those randomized subjects who discontinued the study before this visit. For the EOT visit, the percentage of participants with non-expected visit will be distinguished between those who are still on treatment and those who are not and never had the chance to complete the questionnaires.

A questionnaire is considered returned when it has been received with a completion date. In order to return the questionnaire, all items must be answered and therefore all scores can be calculated. See Section 14.2.4 and 14.2.5 for scoring details.

10.2 Important Protocol Deviations

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a participant's rights, safety, or well-being.

Important protocol deviations include:

- Participants enrolled and dosed on the study who did not satisfy enrolment criteria
- Participants that develop withdrawal criteria whilst on the study but are not withdrawn
- Participants that receive the wrong study intervention or an incorrect dose
- Participants that receive an excluded concomitant medication
- Failure to collect data necessary to interpret primary endpoints
- Failure to collect necessary key safety data
- Any other protocol deviation that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a participant's rights, safety, or well-being.

Important protocol deviations will be identified for all participants by either site monitoring, medical review processes or programming and confirmed prior to or at the Cross-functional Protocol Deviation Review Meetings.

All protocol deviations will be documented in SDTM datasets whether identified through site monitoring, medical review or programming. The management of protocol deviations is outside of this IAP document.

Important protocol deviations or important events that might have an effect on PK include, but may not be limited to the following:

- Sample processing errors that may lead to inaccurate bioanalytical results
- Inaccurate dosing or dosing errors (e.g., dose administration delayed, dose change or missed doses)
- Pre-dose or trough sample collected after the actual dosing
- Concomitant medication, vitamins, dietary or herbal supplements that could affect PK

Should one or more of these events be available at the Data Review Meeting, its implication for PK evaluation will be discussed and agreed amongst relevant study team members (e.g., Sponsor Clinical Pharmacology/Biostatistics/Clinical Pharmacokinetics & CCI team representative). Appropriate action will be taken such as flagging individual values to be excluded from analysis.

A frequency table as well as a listing of important protocol deviations will be provided. In the table, frequencies will be presented by study intervention group and total. The number and percentage of participants by protocol deviation categories defined in the Protocol Deviation Management Plan will be summarized.

In case the JB100 data is declared suitable, a separate frequency table will be provided for all JB100 participants in the avelumab + BSC arm (n=350). Only inclusion and exclusion criteria not met with respect to the Medley protocol (i.e., not with respect to the original JB100 protocol) identified by programming or medical review, if necessary, will be described. For programming purposes, the inclusion and exclusion criteria of JBM study are compared and matched with those from the JB100 study as described in [Appendix 2](#). The number and percentage of participants by inclusion and exclusion category will be presented.

11 Demographics and Other Baseline Characteristics

11.1 Demographics

Demographic characteristics will be summarized descriptively using the following information from the “Demographics” eCRF page.

The following demographic characteristics will be included:

- Sex: male, female, undifferentiated
- Race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, More than one race (combinations will also be displayed if any), Other, Not permitted per local regulation, for JB100: NA in France, Not applicable, Not reported
- Ethnicity: Hispanic or Latino, Not Hispanic or Latino, Not permitted per local regulation, Not reported
- Age at informed consent (years)
- Age categories: < 65 years, ≥ 65 years, 65-74, 75-84, ≥ 85 years

- Pooled Region: North America, Europe, Asia, Rest of the World
 - Geographic Region: North America, Latin America, Western Europe, Eastern Europe, Middle East, Australasia, Asia
- Site codes will be used for the determination of the participant's geographic region. Only geographic regions with at least one participant will be displayed.
- European Economic Area (EEA): Yes, No

A listing containing the demographic information will be produced; the following data collected at baseline will also be presented:

- Height (cm)
- Weight (kg)
- BMI (kg/m²)
- BSA (m²)
- ECOG Performance Status: 0, 1, 2, 3, 4
- Creatinine clearance (see Section 15.3)

Specifications for computation:

- $BSA [m^2] = \sqrt{\frac{height[cm] \times weight[kg]}{3600}}$
- $BMI [kg/m^2] = \frac{weight [kg]}{height[cm]^2} \times 10000$

In the formula for BSA and BMI, values at baseline should be used for both weight and height.

11.2 Medical History

The medical history will be summarized from the “Medical History” eCRF page, using the most recent MedDRA version at time of analysis, preferred term as event category and System Organ Class (SOC) body term as Body System category. Each participant will be counted only once within each Preferred Term (PT) or SOC.

Medical history will be displayed in terms of frequency tables ordered by primary SOC in alphabetic order and PT in descending frequency of the overall total column.

A listing will be provided with SOC term, PT term, start date, toxicity grade (if ongoing) and whether related to the study condition.

11.3 Other Baseline Characteristics

11.3.1 Disease History

Information on disease history is collected during the screening visit on the “Disease History” eCRF page. Information on the presence of visceral metastases will be available from the randomization report provided by the IWRS vendor. The ECOG Performance Status will be collected in the “ECOG Performance Status” eCRF page. The CCI status will be obtained from vendor’s data. Data will be summarized as follows:

- Site of primary tumor: renal pelvis, ureter, bladder, urethra, prostate gland
- Time since initial cancer diagnosis (months) = (date of randomization – date of initial cancer diagnosis) / 30.4375
- Time since documented metastatic disease diagnosis (months) = (date of randomization – date of documented metastatic disease diagnosis) / 30.4375
- Histopathological classification: transitional cell, mixed transitional/non-transitional cell
- TNM classification at initial diagnosis
- TNM classification at study entry
- Clinical stage at initial diagnosis
- Clinical stage at study entry
- Liver lesions at study entry: yes, no
- Lung lesions at study entry: yes, no
- Sites of metastases prior to start of first line platinum regimen (e.g.: lung, liver, colon, etc.)
- Visceral metastases: present, absent (as indicated at the time of the randomization in the IWRS system)
- ECOG Performance Status: 0, 1 and any higher value if present in the data
- CCI

CCI status at baseline is provided by results of the immunohistochemistry analysis done on tumor samples collected at Screening.

CCI status is considered positive if any of the following criteria is met:

- a. Tumor cells expressing CCI $\geq 25\%$
- b. Immune Cells Present $> 1\%$ and Immune cells expressing CCI $\geq 25\%$
- c. Immune Cells Present = 1% and Immune cells expressing CCI = 100%

Otherwise, if results are available but none of the criteria above is met, CCI status will be considered negative. If no results are available, CCI will be left missing.

Incomplete dates for initial cancer diagnosis and documented metastatic disease diagnosis will be handled as specified in Section 9.9.

Analysis: PA

A listing presenting the data collected on the “Disease History” eCRF page will be produced. In addition, a separate listing will present the **CCI** status at Screening.

11.3.2 Nicotine Usage

Analysis: PA

Information on nicotine usage is collected during the screening visit on the “Nicotine Usage” eCRF page.

The following will be tabulated:

- Usage status for each type of nicotine product: never, former, current
- Smoking exposure for cigarettes, cigars and pipes use (pack-years): 0, < 20, 20 – < 40, ≥ 40 and summary statistics. For the calculation of the number of pack-years see Section 9.5.

A listing presenting all the data collected in the eCRF page will be presented.

11.4 Prior Anticancer Therapy

Analysis: PA

The prior anti-cancer therapies are collected under the “Prior Anti-Cancer Drug Therapies Details”, “Prior Anti-Cancer Radiotherapy Details” and “Prior Anti-Cancer Surgeries Details” eCRF pages.

The number and percentage of participants in each of the following anticancer therapy categories will be tabulated:

- Prior anticancer drug therapy for locally advanced or metastatic disease (i.e., one line of platinum-containing chemotherapy)
 - Any prior anticancer drug therapy: Yes / No
 - Prior anticancer drug therapy
 - Duration of prior anticancer drug therapy (weeks): summary statistics. To be calculated as (end date – start date + 21) / 7. The start and end dates correspond to the earliest start date and latest end date among all the drugs administered during the therapy, respectively.
 - Best response to prior anticancer drug therapy: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Non-Complete

Response/Non-Progressive Disease (Non-CR/Non-PD) / Not Evaluable / Unknown
/ No target lesion at baseline

- Time from last dose of prior anticancer drug therapy to first dose of study intervention (weeks): summary statistics. To be calculated as (date of first dose of study intervention – date of last dose of prior therapy) / 7. The date of last dose is the latest end date among all the drugs administered during the therapy.
- Any prior anticancer radiotherapy: Yes / No
- Any prior anticancer surgery: Yes / No

Listings of prior anticancer drug therapies, radiotherapy and surgeries will present the data as collected on the corresponding eCRF pages.

12 Previous or Concomitant Therapies/Procedures

12.1.1 Previous and Concomitant Medications

Analysis: PA

Concomitant medications are medications, other than study intervention and premedications for study interventions, which are taken by participants any time during the on-treatment period, see Section 9.7.

Previous medications are medications, other than study intervention and premedications for study interventions, which started on or before first administration of study interventions.

A medication may be classified as both concomitant and previous. The respective flags will be derived based on start and end date. To derive these flags in case of missing or partial dates, see Section 9.9.

Concomitant and previous medications will be summarized by number and percentage of participants from the “Concomitant Medications Details” eCRF page. Preferred term (PT) will be tabulated as given in the most current version of the WHO-DD dictionary at the time of analysis. The summary tables will be sorted by PT in descending frequency of the overall total column. In case of equal frequency regarding drug class or drug name, alphabetical order will be used. In case any specific medication does not have a PT, it will be summarized under the “UNCODED” category. Each participant will only be counted once, even if he/she received the same medication at different times.

Any medication recorded on the “Concomitant Medications Details” eCRF page will be listed with an indication of whether the medication was previous, concomitant or both.

12.1.2 Premedication

Analysis: PA

Premedication is expected to be administered on the same day as, but prior to, the study intervention administration following the instructions provided in the protocol. Information related to premedication is collected in the “Premedication Details” eCRF page.

Drugs administered for premedication will be summarized by PT in the same way as the concomitant and previous medications (see Section 12.1.1). Premedication data will also be listed.

12.1.3 Previous and Concomitant Procedures

Analysis: PA

Procedures performed during the study will be collected on the “Concomitant Procedures Details” eCRF page. Concomitant procedures will be classified using the most recent MedDRA version at the time of analysis.

All **concomitant procedures** which were undertaken during the on-treatment period will be summarized. Concomitant procedures will be presented according to MedDRA SOC and PT.

A listing presenting the information about concomitant procedures will be produced. A flag will be added to indicate whether the procedure is:

- Prior: the procedure started before the first dose of study intervention
- Concurrent: the procedure started or was ongoing during the on-treatment period.
- Post: the procedure started after the on-treatment period.

12.1.4 Subsequent Anticancer Therapy

Analysis: PA

The subsequent anticancer therapies are collected under the “Anti-Cancer Treatment After Discontinuation Details”, “Radiotherapy after Discontinuation Details” and “Surgery after Discontinuation Details” eCRF pages.

The number and percentage of participants in each of the following anticancer treatment categories will be tabulated:

- Any subsequent anticancer treatment: Yes / No
- Any subsequent anticancer drug therapy: Yes / No
- Number of subsequent anticancer drug therapy regimens: 1 / 2 / 3 / ≥ 4
- Best response across all subsequent anticancer drug therapies: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Non-Complete Response/Non-Progressive Disease (Non-CR/Non-PD) / Not Evaluable / Unknown.
- Any immune check-point inhibitor:

- Pembrolizumab
 - Pembrolizumab + Enfortumab Vedotin
 - Nivolumab
 - Nivolumab + Ipilimumab
 - Nivolumab + Cisplatin
 - Nivolumab + Cisplatin + Gemcitabine
 - Any additional immune check-point inhibitor that is approved during this study may also be added to the list
- Any subsequent anticancer drug therapy of interest:
 - Antibody Drug Conjugates: Enfortumab Vedotin, Sacituzumab Govitecan
 - Fibroblast growth factor receptor inhibitor: Erdafitinib
 - Small molecules of tyrosine kinase inhibitors: Cabozantinib
 - Chemotherapy: Vinflunine, Paclitaxel, Nab-paclitaxel, Gemcitabine, Cisplatin, Carboplatin, Doxorubicin, Vinblastine
 - Vascular endothelial growth factor: Ramucirumab, Bevacizumab
 - Antifolates: Pemetrexed
 - PARP inhibitors: Olaparib, Niraparib
- Any subsequent anticancer radiotherapy: Yes / No
- Any subsequent anticancer surgery: Yes / No

Subsequent anticancer drugs will be summarized by PT in the same way as the concomitant and previous medications (see Section 12.1.1).

The listings of subsequent anticancer treatments will also be provided as follows. These will include all the data collected on the corresponding eCRF pages.

- Listing of subsequent anticancer drug therapies
- Listing of subsequent anticancer radiotherapy and surgery

13 Study Intervention: Compliance and Exposure

Note: If JB100 data is declared suitable, all analyses described under “avelumab” will be also performed for the external data.

The administration schedule for each study intervention group is provided in Section 6. According to the study protocol, in the combination groups, avelumab should be infused before SG and M6223, but after NKTR-255 (for more details see Section 1.3 of the study protocol). All dosing calculations and summaries will be based on “Avelumab Administration Details”, “Sacituzumab

Govitecan Administration Details”, “M6223 Administration Details” and “NKTR-255 Administration Details” eCRFs pages.

13.1 Dose Adjustments

Infusion Planned and Actual Dose

For each infusion, the planned dose for SG and NKTR-255 will be indicated in the eCRF as it is based on the weight of the participant; the planned dose for avelumab and M6223 will be the one indicated in the protocol (i.e., 800mg for avelumab and 1600mg for M6223).

The actual dose administered for each infusion will be collected in the eCRF for all study drugs.

SG is the only drug in this study where a dose reduction can be performed, i.e., the planned dose for an infusion may be reduced compared to the dose that should be administered per protocol. These reductions will be referenced to as planned dose adjustments. The investigator will decide for how long the planned dose adjustment should be applied. For avelumab, M6223 and NKTR-255, planned dose adjustments are not allowed.

For all four drugs, at each visit, the planned dose to be infused may be higher than the actual dose infused due to infusion-related reactions or other reasons that lead to the permanent interruption of the infusion at that visit. These cases will be referenced to as unplanned dose adjustments.

For SG, a planned dose adjustment will be differentiated from an unplanned dose adjustment when both of the following criteria are met:

- The planned dose is the same as the actual dose
- The answer to the question “Is there a change in dose?” is “Dose adjusted”

Minimum Dose Level

For each infusion of each drug, the percentage of the planned dose actually administered will be calculated for all the unplanned dose adjustments. Then, for each participant, the minimum percentage for each applicable study drug will be selected.

The following information related to dose adjustments will be summarized:

For avelumab:

- Number and percentage of participants who did not receive avelumab
- Number and percentage of participants without any unplanned dose adjustment for avelumab
- Number and percentage of participants with at least one unplanned dose adjustment for avelumab and number of unplanned dose adjustments (1, 2, 3, ≥ 4)
- Number and percentage of participants by minimum unplanned avelumab dose level: $< 50\%$, $50 - < 70\%$, $70 - < 90\%$, $90 - < 100\%$

- Number and percentage of participants with at least one unplanned dose adjustment for avelumab due to AE or other reason

For the combination drug (i.e., SG, M6223, NKTR-255):

- Number and percentage of participants who did not receive the combination drug
- Number and percentage of participants without any unplanned dose adjustment for the combination drug
- Number and percentage of participants with at least one unplanned dose adjustment for the combination drug and number of unplanned dose adjustments (1, 2, 3, ≥ 4)
- Number and percentage of participants by minimum unplanned dose level for the combination drug: < 50%, 50 - < 70%, 70 - < 90%, 90 - < 100%
- Number and percentage of participants with at least one unplanned dose adjustment for the combination drug due to AE or other reason

For SG only:

- Number and percentage of participants without planned dose adjustment
- Number and percentage of participants with at least one planned dose adjustment
- Number and percentage of participants with at least one planned dose adjustment due to AE or other reason

13.2 Therapy Delays

Delays will be derived individually on each drug and based on the actual infusion dates. Delays of up to two days are allowed per protocol and will not be considered as delays. For example, if one participant receives avelumab on day 1, then the next avelumab administration date is planned for day 15; however, if the participant receives avelumab at day 16 or 17, this is considered as ‘no delay’.

The days of delay between two performed infusions (i.e., the actual dose is not missing and greater than zero) will be calculated as follows:

- Avelumab and M6226: date of infusion – date of previous infusion – 14
- SG:
 - If the previous infusion was a Day 1 infusion of a cycle: date of infusion – date of previous infusion – 7
 - If the previous infusion was a Day 8 infusion of a cycle: date of infusion – date of previous infusion – 14
- NKTR-255: date of infusion – date of previous infusion – 28

The following information on study intervention delays will be tabulated:

- Number and percentage of participants with and without infusion delays
- Number and percentage of participants by maximum length delay: no delay (including 1-2 days delays), 3 - 6 days, 7 or more days
- Number and percentage of participants by reason for no dose as collected in the eCRF pages

13.3 Exposure

The following parameters will be provided in summaries and listings for each study drug separately:

- Exposure duration
- Total number of infusions received
- Cumulative actual dose
- Dose intensity
- Relative dose intensity

For the purpose of the derivation of the quantities described below, the duration of a cycle is as follows:

- Avelumab and M6223: 2 weeks
- SG: 3 weeks
- NKTR255: 4 weeks

Exposure Duration

Summary statistics for the exposure duration to study intervention and to each study drug will be tabulated. See Section 9.8 for the derivation of the exposure duration for each study intervention group as well as for each study drug.

Analysis: IA

A Kaplan-Meier plot will be produced to present the number of participants on treatment by the time since first dose (months). The exposure end date of study intervention is calculated as indicated in Section 9.8 and will be used as date of event for those participants who discontinued study intervention or as censoring date for those who are still on treatment at the time of analysis.

Number of Infusions

Analysis: PA

Summary statistics for the total number of infusions received will be provided for each study drug. Only infusions with an actual dose greater than zero will be considered.

Cumulative Dose

Analysis: PA

The cumulative dose per participant is the sum of the actual doses that the participant received while on-treatment. It will be measured in mg for avelumab and M6223, mg/kg for SG and avelumab in JB100, µg/kg for NKTR-255. Summary statistics will be provided.

Dose Intensity

The dose intensity will be calculated as follows:

- Avelumab and M6223 (mg/cycle) = $\frac{\text{total cumulative dose (mg)}}{\text{exposure duration (in weeks)}/2}$
- Avelumab in JB100 (mg/kg/cycle) = $\frac{\text{total cumulative dose (mg/kg)}}{\text{exposure duration (in weeks)}/2}$
- SG (mg/kg/cycle) = $\frac{\text{total cumulative dose (mg/kg)}}{\text{exposure duration (in weeks)}/3}$
- NKTR-255 (µg/kg/cycle) = $\frac{\text{total cumulative dose (µg/kg)}}{\text{exposure duration (in weeks)}/4}$

Summary statistics will be provided.

Relative Dose Intensity

The relative dose intensity will be calculated as follows:

- Avelumab (%) = $100 \times \frac{\text{dose intensity (mg/cycle)}}{800 \text{ mg/cycle}}$
- Avelumab in JB100 (%) = $100 \times \frac{\text{dose intensity (mg/kg/cycle)}}{10 \text{ mg/kg/cycle}}$
- SG (%) = $100 \times \frac{\text{dose intensity (mg/kg/cycle)}}{20 \text{ mg/kg/cycle}}$
- M6223 (%) = $100 \times \frac{\text{dose intensity (mg/cycle)}}{1600 \text{ mg/cycle}}$
- NKTR-255 (%) = $100 \times \frac{\text{dose intensity (µg/kg/cycle)}}{3 \text{ µg/kg/cycle}}$

Summary statistics will be provided. In addition, the number and percentage of participants in the following categories will be presented: < 80%, 80% - < 90%, 90% - 110%, > 110%.

Analysis: PA

A listing presenting the data collected on “Avelumab Administration Details”, “Sacituzumab Govitecan Administration Details”, “M6223 Administration Details” and “NKTR-255 Administration Details” eCRFs pages will be produced. An additional listing will also be created to present the exposure duration, number of infusions, cumulative dose, dose intensity and relative dose intensity described above.

14 Efficacy Analyses

The following analyses will focus on the following comparisons:

- Group B vs. Control Group
- Group C vs. Control Group
- Group D vs. Control Group

Control Group may refer to either Group A participants only or the extended control group including the data from JB100 together with Group A participants. For JB100, only data from investigator tumor assessments will be used.

The information collected in the “Tumor Assessment (According to RECIST 1.1) - Target Lesions”, “Tumor Assessment (According to RECIST 1.1) - Non-target Lesions”, “Tumor Assessment (According to RECIST 1.1) - New Lesions”, “Sum of Diameters (According to RECIST 1.1) and “Assessment of Disease Based on Imaging (According to RECIST 1.1)” eCRF pages will be listed.

Analysis: IA

If JB100 data are considered suitable, all the analyses described in Section 14.1.1, with the exception of the verification of the model assumptions and the restricted mean survival time analysis, will be repeated excluding JB100 data.

Analysis: PA

If JB100 data are considered suitable, all the analyses described in Section 14.1 and Section 14.2.1 will be repeated excluding JB100 data.

14.1 Primary Estimand: PFS

For details about the estimand attributes, please refer to Section 5.

14.1.1 Primary Objective: Derivation and Analysis of the Primary Estimand PFS

PFS time is defined as the time from date of randomization to PD according to RECIST v1.1 as assessed by Investigator or death due to any cause, occurring within two scheduled tumor assessments after last evaluable assessment or randomization.

PFS time will be computed as follows:

$$\text{PFS time (months)} = (\text{date of PD or death or censoring} - \text{date of randomization} + 1) / 30.4375$$

The interval of two scheduled tumor assessments is defined as follows:

- 112 days (16 weeks) during the first year after randomization
- 168 days (24 weeks) after the first year from randomization

The last adequate tumor assessment is defined as the last tumor assessment with overall response not equal to “NE”.

If a tumor assessment was performed on the same day as the start of subsequent anticancer treatment, it will be assumed that the tumor assessment was performed prior to the start of the new anticancer therapy and therefore it will be included in the derivation of the PFS time.

The censoring and event date options to be considered for the PFS analysis are presented in Table 4.

Table 4 Outcome and Event Dates for PFS Analysis

Scenario	Date of event/censoring	Outcome
No baseline or post-baseline assessment	Randomization date	Censored ^a
PD or death ≤ interval of two scheduled tumor assessments after last adequate tumor assessment or after randomization	Date of PD or death, whichever is earlier	Event
PD or death > interval of two scheduled tumor assessments after the last adequate tumor assessment	Date of last adequate tumor assessment ^b documenting no PD before subsequent anticancer treatment is given or missed tumor assessments	Censored
No PD nor death	Date of last adequate tumor assessment ^b documenting no PD before subsequent anticancer treatment is given or missed tumor assessments	Censored
Discontinuation of study intervention due to ‘Disease progression’ without documented progression	Not applicable	Information is ignored. Outcome is derived based on documented progression only.
Subsequent anticancer therapy given before PD or death event	Date of last adequate tumor assessment ^b documenting no PD before subsequent anticancer treatment is given or missed tumor assessments	Censored

^a However if the participant dies ≤ 16 weeks after start date the death is an event with date on death date.

^b If there are no adequate post-baseline assessments prior to the PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria were met the censoring will be on the date of randomization.

Frequency (number and percentage) of participants with each event type (PD or death) and censoring reasons will be presented by intervention group. Reasons for censoring will be summarized according to the categories in Table 5. The PFS time or censoring time and the reasons for censoring will also be listed.

Table 5 Censoring Reason

Hierarchy	Condition	Censoring Reason
1	No event and none of the conditions in the below hierarchy are met	Ongoing in the study without an event
2	No baseline tumor assessment	No baseline tumor assessment
3	No event and no evaluable post-baseline tumor assessment	No post-baseline tumor assessment
4	Start of subsequent anticancer treatment before event	Start of subsequent anticancer treatment
5	Event after the interval of two scheduled tumor assessments from the last evaluable tumor assessment if available, or from randomization	Event after 2 or more missing assessments
6	No event and “Withdrawal by subject” in study termination eCRF pages	Withdrawal of consent
7	No event and - lost to follow-up in any treatment termination or study termination eCRF pages - last date known to be alive > 13 weeks prior to the analysis cut-off	Lost to follow-up

Stratified Logrank Test and Cox’s Regression Model

The PFS time will be compared between each combination group and control group using a two-sided stratified logrank test (test=logrank option in SAS PROC LIFETEST).

The study intervention effect between each combination group and control group will be evaluated in terms of the hazard ratio and 95% CI, estimated by means of a Cox proportional hazards model stratified by presence of visceral metastases at baseline (yes/no). Randomization strata will be taken as specified and documented in the IWRS and each stratum will define a separate baseline hazard function (using the “STRATA” statement in SAS PROC PHREG), i.e., for the i-th stratum, the hazard function is expressed as: $h(i;t) = h(i,0;t) \exp(\beta X)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum, x defines the intervention group (0 = Control Group, 1 = Combination Group), and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model with the discrete logistic model (Ties=Discrete option in SAS PROC PHREG). With this option, the score test in the Cox proportional hazards model is identical to the logrank test.

In case JB100 data will be included in the main analysis, propensity score-based weights as described in Section 9.10.5 will be used for the logrank test and in the Cox proportional hazards model (weight = w option in SAS PROC LIFETEST and SAS PROC PHREG).

In this case, ties in the Cox proportional hazards model will be handled by the Efron approximation (Ties = Efron option in SAS PROC PHREG).

Kaplan-Meier Estimates

Kaplan-Meier product-limit estimates will be presented together with a summary of associated statistics for PFS: median and corresponding 95% CI, Q1 and Q3, minimum and maximum. Progression-free rate estimates at 3, 6, 9, 12, 15, 18 months and for every 3 months thereafter as applicable, including the corresponding two-sided 95% CIs, will also be reported. The CI for the median will be calculated according to Brookmeyer and Crowley (1982) and CIs for the survival function estimates at the above defined time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (1980) (conftype=loglog default option in SAS PROC LIFETEST). The estimate of the standard error will be computed using Greenwood's formula. A Kaplan-Meier plot will also be produced.

In case JB100 data will be included in the main analysis, propensity score-based weights as described in Section 9.10.5 will be used to calculate Kaplan-Meier estimates (weight = w option in SAS PROC LIFETEST).

A table including the Kaplan-Meier estimates, the stratified Cox proportional hazards model results as well as the p-value of the stratified logrank test will be provided.

Methods for Evaluating the Validity of Model Assumptions

Following visual checks will be done to evaluate the validity of proportional hazards assumed for the Cox model as described above:

1. In case JB100 data will be included in the main analysis, the proportional hazards assumption will be checked visually by presenting weight-adjusted Kaplan-Meier curves including the investigational arm, the randomized control arm, the external control arm and the overall control group combining randomized and external control. Same weights as calculated for the main analysis will be used. Crossing curves indicate a violation of the proportional hazards assumption. The same visual check will be done in case the external data is not used. A Kaplan-Meier plot for this check is already available as part of the main analysis (see Section above).
2. Another check will be done by plotting

$$\log(-\log(S(t))) \text{ versus } \log(t)$$

for each randomization stratum. $S(t)$ is the estimated survival function for PFS at time t by treatment arm (plots=lls) in SAS PROC LIFETEST) based on the main analysis with or without weights, depending on the comparability of randomized and external data. The assumption of proportional hazards holds true when the curves are reasonably parallel.

3. Schoenfeld residuals for the stratified Cox proportional hazard model including a LOESS curve will be plotted; a non-zero slope is evidence of deviation from proportional hazards.

Schoenfeld residuals will be computed in SAS using the PHREG procedure and using the OUTPUT statement and the keyword RESSCH. With proportional hazards the LOESS curve should be parallel to the x-axis. The LOESS curve will be plotted using the SAS SGPLOT procedure with the LOESS statement and the CLM option to add the 95% confidence limits. Weights will be added to the procedure in case JB100 data is included in the main analysis.

Besides the above described analyses of the primary estimand and evaluation of the validity of certain model assumptions, PFS will also be analyzed based on restricted mean survival time (RMST) differences (Zhang, 2013).

Restricted Mean Survival Time (RMST)

The RMST methodology is applicable independently of the proportional hazards assumption. The RMST is defined as the area under the survival curve up to time τ . The cut-off point τ should not exceed the minimum of the largest observed time for the compared intervention groups so that the RMST of all intervention groups being evaluated can be adequately estimated and comparison between groups is feasible. τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval. The RMST up to time τ can then be interpreted as the expected PFS time restricted to the same follow-up time τ among all participants. The selection of τ should ensure that the RMST evaluation will not go beyond the maximum time point where the evaluation can be performed while also taking into account a large period of time that is expected to provide a meaningful assessment of study intervention effect. To avoid arbitrary selection of the common cut-off τ for both intervention groups, three sets of analyses will be performed:

- τ_1 = minimum of (largest observed PFS time for control group, largest observed PFS time for combination group).
- τ_2 = minimum of (largest PFS event time for control group, largest PFS event time for combination group).
- τ_3 = midpoint between τ_1 and τ_2 .

Regardless of the usage of external data, τ will be determined based on randomized data only. To calculate the area under the curve, Kaplan-Meier estimates are calculated as described above, i.e., in case JB100 data is included in the main analysis, weights will be included to create PFS curves.

The study intervention effect between the control group and the combination group will be assessed based on the difference of the areas under the curve over the respective pre-specified time interval. The associated 95% CI for the difference in means and two-sided p-value will be generated.

14.1.2 Sensitivity Analyses of the Primary Estimand PFS

Sensitivity 1

To evaluate the robustness of the treatment effect with regard to stratification, the same analysis as described in Section 14.1 will be performed but without the stratification factor in the Cox model.

Sensitivity 2

To assess the heterogeneity of study intervention effects across the subgroup levels defined in Section 8.2, two Cox models will be fitted with PFS time as the dependent variable; subgroup, treatment, and with and without the treatment-by-subgroup interaction as explanatory variables.

- Model 1: treatment + subgroup
- Model 2: treatment + subgroup + treatment $\times$ subgroup

A p-value for the interaction test (Likelihood Ratio test) will be provided together with the HR and corresponding 95% CI of the interaction model parameter.

Sensitivity 3

To explore the homogeneity of the study intervention effect across subgroups as defined in Section 8.2, for each subgroup level, a Cox model will be fitted with PFS time as dependent variable and treatment as only covariate.

The HR and its corresponding 95% CI of all subgroup levels will also be presented in a forest plot.

Sensitivity 4

In order to evaluate the robustness of the estimand towards the assumption that tumor assessments after several missing scans cannot reasonably be used to determine the time point of progression, the analysis described in Section 14.1.1 will be repeated considering all RECIST PDs and deaths as event, i.e., events after two or more missing tumor assessments are included.

Sensitivity 5

Analysis: PA

In order to evaluate the impact of selecting the wrong category of the stratification factor at the time of randomization, the same analysis as described in Section 14.1 and the Sensitivity 1 analysis described in this section will be repeated considering the stratification factor indicated in the eCRF.

Sensitivity 6

Analysis: PA

If JB100 data are considered suitable, all the analyses described in Section 14.1.1, with the exception of the verification of the model assumptions and the restricted mean survival time analysis, will be repeated using the propensity scores obtained by removing the option “MISSING” from the PROC LOGISTIC (see Section 9.10.3). Therefore, JB100 participants with at least one

baseline covariate missing will not have the propensity score derived and will be excluded from the analysis where propensity score-based weights are used.

14.1.3 Supplementary Analyses of the Primary Estimand PFS

Analysis: PA

To evaluate if the study intervention improves PFS irrespective of subsequent anticancer treatment administered before PD, the analysis described in Section 14.1.1 will be repeated considering also tumor assessments after the start of subsequent anticancer treatment.

14.1.4 Follow-up Duration of PFS

Note: In case JB100 data is included in the main analysis, weights will be included for Kaplan-Meier estimates and the plot.

A Kaplan-Meier plot for follow-up duration of PFS will be generated to assess the follow-up time in the intervention groups reversing the PFS censoring and event indicators. Kaplan-Meier estimates (product-limit estimates) will be presented by intervention group together with a summary of associated statistics including the median time of follow-up for PFS with 2-sided 95% CIs. In particular, the rates at 3, 6, 9, 12, 15, 18 months and for every 3 months thereafter as applicable will be estimated with corresponding 2-sided 95% CIs.

14.2 Further Estimands

For details about the estimand attributes, please refer to Section 5.

14.2.1 Secondary Estimand: OS

Note: In case JB100 data is suitable and used for the PFS analysis, the same applies to the OS analysis.

OS time is defined as the time from date of randomization to death due to any cause.

OS time will be computed as follows:

$$\text{OS time (months)} = (\text{date of death or censoring} - \text{date of randomization} + 1) / 30.4375$$

For participants not known to be deceased at time of analysis or who died after the cut-off date, OS time will be censored at the date of last contact before data cut-off date as defined in Section 9.6.

Frequency (number and percentage) of participants with an event (death) and censoring reasons will be presented by study intervention group. Censoring reasons are as follows:

- Alive
- Withdrawal of consent

- Lost to follow-up

Lost to follow-up will include participants with a last contact date > 13 weeks prior to the analysis cut-off date, as well as participants that, according to the investigator, were lost to follow-up prior to the analysis cut-off (information available from treatment termination or study termination eCRF pages).

The OS time or censoring time and the reasons for censoring will also be listed.

The analyses described in Section 14.1 for PFS will be repeated for OS except for sensitivity analysis 4 and 5, and the supplementary analysis.

14.2.2 Secondary Estimand: OR

Note: Since the propensity score-based weights are built on pre-chosen baseline covariates with possible impact on the PFS and/or OS outcome but independent on their impact on objective response, in the following, statistical calculations, e.g., Cochran-Mantel-Haenszel test (CMH), will not include (weighted) data from JB100. However, descriptive analyses, e.g., number and percentages, will show JB100 data in a separate column in case the external data was declared suitable.

Best Overall response (BOR) will be assessed based on reported overall lesions responses at different evaluation time points from the date of randomization until first documentation of PD in accordance with RECIST v1.1, taking requirements for confirmation into account as detailed below. Clinical deterioration will not be considered as documented disease progression. Only tumor assessments performed on or before the date of start of subsequent anticancer treatment will be considered in the assessment of BOR.

The date of the overall response assessment is the earliest date for imaging of target, non-target and new lesions of images taken at that response assessment. No imputation is planned for these dates as complete dates are required for these assessments.

BOR Based on Confirmed Responses:

- CR = at least two determinations of CR at least 4 weeks apart (with no PD in between)
- PR (applicable only to participants with measurable disease at baseline) = at least two determinations of PR or better (PR followed by PR or PR followed by CR) at least 4 weeks apart (and not qualifying for a CR), with no PD in between
- SD (applicable only to participants with measurable disease at baseline) = at least one SD assessment (or better) $\geq$ 6 weeks after the date of randomization (and not qualifying for CR or PR)
- Non-CR/non-PD (applicable only to participants with non-measurable disease at baseline) = at least one non-CR/non-PD assessment (or better) $\geq$ 6 weeks after the date of randomization (and not qualifying for CR).

- PD = PD $\leq$ 12 weeks after the date of randomization (and not qualifying for CR, PR, SD or non-CR/non-PD)
- Not Evaluable (NE): all other cases.

SD can follow PR only in the rare case that tumor increases by less than 20% from the nadir, but enough that a previously documented 30% decrease from baseline no longer holds. If this occurs, the sequence PR-SD-PR is considered a confirmed PR. A sequence of PR-SD-SD-PD would be a best response of SD if the minimum duration for SD definition has been met.

Participants without target nor non-target lesions at baseline may have subsequent tumor assessments which show no evidence of disease (i.e., no new lesions). In such cases, sites have been asked to indicate CR as overall response if the tumor assessment is a complete assessment (i.e., all the body areas that are to be monitored have been scanned) or NE if it is a partial assessment (i.e., only for some body areas the scans are available). Those participants with an overall response of CR will be assigned an overall response of “No Evidence of Disease” (NED) at the time point response. However, since an overall response of NED is not a valid response for use for the derivation of BOR, these participants will have a BOR of PD if progressive disease occurred within 12 weeks from the date of randomization, otherwise BOR will be NE.

Objective Response

Participants are defined as having an Objective Response (OR) if they achieved either a confirmed BOR of CR or PR according to RECIST v1.1 from randomization to first observation of PD.

The OR rate is the proportion of participants with OR.

Disease Control

Participants are identified as having objective disease control (DC) if they achieve a response of either a confirmed CR or PR, SD or non-CR/non-PD, as defined above.

DC rate is the proportion of participants with DC.

The number and percentage of participants with BOR of confirmed CR, confirmed PR, SD, non-CR/non-PD, PD, and NE as well as the number of participants achieving objective response and disease control, the OR and DC rates and the corresponding 2-sided exact Clopper-Pearson 95% CI (BINOMIAL (EXACT) option of the TABLES statement in the SAS PROC FREQ) will be presented.

The association of study intervention (considering the investigational combination group and Group A) and objective response will be tested by the General Association Statistic of the Cochran-Mantel-Haenszel test (CMH) with the randomization strata taken into account. The null hypothesis of no association in any of the randomization strata is tested against the alternative, which specifies that there is an association between study intervention effect and tumor response at least in one randomization stratum. The CMH test will be performed at two-sided alpha level of 0.05.

The stratified odds ratio in terms of objective response will also be estimated along with its 95% CI to compare the study intervention effect. The odds ratio is defined as the odds of objective response with the combination treatment divided by the odds of objective response with avelumab monotherapy. The Breslow-Day test will be used to check the homogeneity of the odds ratio across the randomization strata. It tests the null hypothesis that odds ratios in all strata are equal against the alternative hypothesis that at least in one stratum the odds ratio is different.

In case the null hypothesis of homogeneity of odds ratios across strata is not rejected at the alpha level of 5% two-sided, the common odds ratio will be determined using the Mantel-Haenszel estimate (by the FREQ procedure using CMH option in SAS); if the null hypothesis of homogeneity of odds ratio across all strata is rejected, the odds ratio per stratum will be calculated with the corresponding exact CI.

Analysis: PA

The unstratified odds ratio in terms of objective response will also be estimated along with its 95% CI for each subgroup (see Section 8.2) and will be presented in a forest plot. In addition, a waterfall plot will show the percent change in sum of target lesion diameters between baseline and the best post-baseline assessment for each participant with measurable disease at baseline and at least one post-baseline assessment where the sum of diameters could be evaluated. Also for this plot, only tumor assessments performed on or before the date of start of subsequent anticancer treatment will be considered. Different colors will illustrate the BOR. If feasible, external data will also be presented in this plot.

The information collected on “Assessment of Disease Based on Imaging (According to RECIST 1.1)” and “Sum of Diameters (According to RECIST 1.1)” as well as the derived BOR and best percent change from baseline (CFB) in terms of sum of diameters will be listed.

14.2.3 Secondary Estimand: DoR

Note: In case JB100 data is suitable, DoR will be given for this data as separate column including frequencies and Kaplan-Meier estimates. No weights will be used.

Duration of Response (DoR) according to RECIST v1.1 assessed by Investigator will only be evaluated for participants with OR (as defined in Section 14.2.2).

DoR is the time from first documentation of OR until PD or death due to any cause, occurring within two scheduled tumor assessments after last evaluable assessment or randomization.

DoR time will be computed as follows:

$$\text{DoR time (months)} = (\text{date of event or censoring} - \text{first date of OR} + 1) / 30.4375$$

DoR time will be censored on the date of the last adequate tumor assessment for participants who do not have an event (PD or death) or start a subsequent anticancer treatment prior to an event or with an event after the interval of two scheduled tumor assessments.

If a tumor assessment was performed on the same day as the start of subsequent anticancer treatment, it will be assumed that the tumor assessment was performed prior to the start of the new anticancer treatment and therefore it will be included in the derivation of the DoR time.

The censoring and event date options to be considered for the DoR analysis are the same as described for PFS (see Table 5).

Frequency (number and percentage) of participants with OR, event and type of event (PD or death) and censored status will be presented by intervention group. Kaplan-Meier estimates described for PFS in Section 14.1.1 will also be calculated for DoR.

The DoR time or censoring time and the reasons for censoring will also be listed.

Analysis: PA

Time on treatment, time to and duration of response for every participant will be plotted using a swimmer plot. Color coded symbols will be used to show the time of the following events: CR, PR, SD, Non-CR/non-PD, PD, NE, death and end of treatment. The color of the lines will indicate the confirmed best overall response.

14.2.4 Secondary Estimand: NCCN FACT FBISI-18 DRS-P

Analysis: PA

Note: Since the propensity score-based weights are built on pre-chosen baseline covariates with possible impact on the PFS and/or OS outcome but independent on their impact on symptoms and quality of life, JB100 data will not be included in the analysis of the DRS-P independent of the suitability.

The NCCN FACT FBISI-18 is designed to be a stand-alone instrument to measure symptoms and quality of life in patients with urothelial carcinoma and was created using inputs from patients and oncologists. The disease-related symptoms - physical (DRS-P) subscale uses a subset of physical symptoms which are specific to urothelial carcinoma, i.e., pain, weight loss, urination, weakness, dizziness, meeting family needs, appetite, erection in males, and sleep. Each question offers 5 answers: not at all (0), a little bit (1), somewhat (2), quite a bit (3), very much (4). The overall summary range is 0-36. The scores of negatively stated items (e.g., pain) are reversed and added together with the positively stated scores. The sum of these item scores are multiplied by the number of items in the subscale, i.e., by 9, and divided by the number of items answered. In general, higher scores are better than lower scores. If at least 50% of items in a subscale have non-missing values, the subscale score is prorated using the method detailed above. See [Appendix 3](#) for the scoring template and [NFBISI-18 \(facit.org\)](#) for further details.

Week 13 assessment is considered a key timepoint for comparative PRO evaluations. This time point would minimize loss of data due to deaths and most participants are expected to be still exposed to treatment. Changes in symptoms as reported by the participants may manifest or likely occur at that time-point.

Two intercurrent events have been identified as of interest for this endpoint: death and treatment discontinuation. The number of deaths and treatment discontinuation at the timepoint of interest will be examined.

Main Estimand

CFB in DRS-P scores will be analyzed using a mixed model for repeated measures (MMRM) analysis using all data from baseline up to 13 weeks (including the W13 visit) for each treatment-control comparison (Group A and B, Group A and C, and Group A and D).

From the MMRM model, least-squares (LS) mean at Week 13 and corresponding 95% CIs will be estimated for each treatment group. The average effect of study interventions at Week 13 will be compared between interventions using the difference in LS means. Two-sided 95% p-value and 95% CI for this difference will be reported. P-values are considered descriptive measures without adjustment for multiplicity.

The MMRM model will include as fixed effect terms: treatment group, visit as categorical variable, treatment-visit-interaction, baseline score and randomization strata (from IWRS). Participant will be included as random effect. Restricted maximum likelihood (REML) estimation will be used. The average effect over visits will be estimated giving each visit equal weight. The unstructured covariance matrix will be used to model the within-subject error. If the model fails to converge a simpler covariance structure will be tried. Denominator degrees of freedom will be computed using Kenward and Roger's method (Kenward 1997).

LS means CFB in DRS-P and respective 95% CIs will be displayed over time. For each treatment-control comparison, a separate figure will be provided.

Sample code using generic visits as categorical variable to be used to analyze each treatment-control comparison:

```
PROC MIXED data=dataset method=reml;
  CLASS subj_id trt visit strata;
  MODEL pro_cfb = trt visit trt*visit pro_base strata /s ddfm=kr;
  REPEATED visit / TYPE=un SUBJECT=subj_id;
  LSMEANS trt*visit / SLICE=visit diff alpha=0.05 cl;
  * estimates treatment by visit interaction using all visits
RUN;
```

with trt as study interventions, visit as generic visit, pro_cfb as CFB DRS-P score, and pro_base as baseline DRS-P score.

The analysis just described will be repeated three times:

- Main analysis: the EOT visit will be excluded from the analysis

- Sensitivity analysis 1: the EOT visit will be analyzed as a separate visit (i.e., no time-windowing will be applied) as long as it happened up to Week 13 visit inclusive (i.e., questionnaires completed within 92 days from first dose).
- Sensitivity analysis 2: as the main analysis but considering for baseline all questionnaires completed up to and including the day of first dose of study intervention, even if collected after the start of the first infusion.

Supplementary Estimand 1

A supplementary estimand will be defined for the CFB in DRS-P score at W13 considering a composite strategy for both intercurrent events: i.e., treatment discontinuation and death will be considered unfavorable outcomes (treatment failure). PROs collected at EOT visit will not be included in the analysis.

Participant's death or treatment discontinuation prior to W13 will result in missing data at the timepoint of interest, i.e., W13. The unobserved data will be assigned a poor outcome, i.e., the worst score DRS-P can take (zero) and the CFB will be calculated again. The completed dataset will then be analyzed via a quantile regression to obtain an estimate of the difference in the median CFB (quantile = 0.5) between the study intervention groups of interest. The model will include the baseline score as a covariate, as well as the randomization strata (from IWRS). Quantile regression will be performed using PROC QUANTREG (SAS procedure).

Sample code:

```
PROC QUANTREG data=dataset;  
  CLASS trt strata;  
  MODEL pro_cfb = trt pro_base strata / quantile = 0.5;  
  TEST trt;  
RUN;
```

with trt as study interventions, pro_cfb as CFB DRS-P score, and pro_base as baseline DRS-P score.

Quantile regression was deemed appropriate for this approach as the assignment of a poor value for data that are missing either due to death or treatment discontinuation is expected to result in a skewed distribution (with a “lump” of observations at one end of the distribution). Quantile regression makes no assumption about the symmetry of the distribution of the CFB, or about the shape of the tails of the distribution. It should be noted that most uses of quantile regression to estimate differences between study intervention groups assume that the difference is solely one of location, i.e., that even if the distribution is not symmetric, the shape of the distribution does not differ significantly between the groups compared. Without this assumption, the estimates of difference from quantile regression are difficult to interpret. Additionally, if the quantile of interest includes the region of the artificial poor score, e.g., if more than half of the participants had died

or discontinued the study treatment before W13, then the use of quantile regression at the median may not be appropriate anymore, as the imputed/assigned values will be part of the estimation process and the resulting treatment effect may not be interpretable.

Supplementary Estimand 2

An additional supplementary estimand will be defined for the CFB in DRS-P score at W13 considering a composite strategy for death, i.e., death will be considered an unfavorable outcome, and a hypothetical strategy for treatment discontinuation. PROs collected at EOT visit will not be included in the analysis.

As mentioned above, participant's death or treatment discontinuation prior to W13 will result in missing data at the timepoint of interest, i.e., W13. The unobserved data due to death prior and up to W13 will be assigned a poor outcome, i.e., the worst score DRS-P can take (zero). All other unobserved data prior and up to W13 will not be imputed and will remain missing. The resulting dataset will then be analyzed via an MMRM model as described in Section 14.2.4 to obtain an estimate of the difference in the mean CFB between the study intervention groups of interest.

CCI

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

CCI

15 Safety Analyses

This section includes specifications for summarizing safety endpoints that are common across clinical studies such as adverse events, laboratory tests and electrocardiogram.

Note: In case the JB100 data is declared suitable for comparison, safety data from the study will be presented in an additional separate column or as part of a figure, e.g., an additional boxplot.

15.1 Adverse Events

Adverse events will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) version (latest version at the time of database lock for each analysis; will be specified in outputs). Severity of AEs will be graded by the investigator using the NCI-CTCAE version 5.0 toxicity grades for JBM data and version 4.03 for JB100 data.

Treatment-emergent adverse events (TEAE) are those events with onset or worsening dates occurring within the on-treatment period as defined in Section 9.7. This includes AEs ongoing at baseline, which first improve under study intervention and then worsen irrespective of baseline. Adverse events with changes in toxicity grade and/or seriousness and outcome of AEs are recorded

as separate entries in the eCRF with associated end and start dates (start date equals end date of previous entry, supported in eCRF by 'AENEWID' in SUPPAE). Records of the same AE will be considered as one event in the analysis. If the toxicity grade of the reported event with onset prior to the first dose of any study intervention worsens during the on-treatment period, the TEAE flag will be re-evaluated for the worse and the subsequent records as per the TEAE definition. If the worse record starts outside of the on-treatment period, it will not appear on the summaries/listings of TEAEs, unless otherwise specified. These events will be kept as separate records in the database in order to maintain the full detailed history of the events. The overall outcome of the adverse event is the outcome of the last event in the sequence. When such AEs are listed, start and end date should be provided together with change date, toxicity grade, outcome, and seriousness per episode. For JB100, each AE entry will be considered as one event as the way AE data were collected does not allow multiple records to be linked to the same AE.

In case AE-related dates are partial or time of AE and/or time of infusion is not complete, the available information will be used in a conservative approach to determine whether the AE is treatment-emergent. Incomplete AE-related dates will be handled as specified in Section 9.9. No imputation will be performed on incomplete or missing time.

If the relationship information is missing for one of the study drugs applicable in the study intervention group to which the participant belongs, the AE will be considered as related to that drug.

All analyses described in this section will be based on TEAEs, if not otherwise specified. According to the type of listing, all AEs or only the TEAEs will be presented. In the TEAEs listings, if an AE becomes treatment-emergent, the whole AE will be listed, and a flag will indicate which records are treatment-emergent.

Unless otherwise specified, TEAEs will be summarized by number and percentage of participants with the TEAE in the category of interest, by study intervention group, primary System Organ Class (SOC) and Preferred Term (PT) in alphabetical order.

Each participant will be counted only once within each SOC or PT. If a participant experiences more than one AE within a SOC or PT for the same summary period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

15.1.1 All Adverse Events

Information related to AEs is collected in the "Adverse Events Details" eCRF page.

Adverse events will be summarized by worst severity per participant, using the latest version of MedDRA PT as event category and MedDRA primary SOC body term as Body System category.

In case a participant has events with missing and non-missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

The overall summary of AEs table will include the frequency (number and percentage) of participants with each of the following:

- TEAEs
- Study intervention related TEAEs (any study intervention and by study drug)
- Serious TEAEs
- Non-serious TEAEs
- Study intervention related serious TEAEs (any study intervention and by study drug)
- TEAEs with NCI-CTCAE Grade ≥ 3 and ≥ 4
- Study intervention related TEAEs with NCI-CTCAE Grade ≥ 3 and ≥ 4 (any study intervention and by study drug)
- TEAEs leading to death
- Study intervention related TEAEs leading to death (any study intervention and by study drug)
- Adverse Events of Special Interest (AESIs):
 - Infusion-related reactions (IRRs)
 - Immune-related AEs (irAEs)
 - Thromboembolic events
 - Cytokine release syndrome
 - QT interval prolongation

AESIs will be identified according to pre-specified search lists based on the last available version of MedDRA Preferred Terms at the time of each analysis. See Section 15.2.3 for more details.

In addition, the frequency of participants with TEAEs by SOC and PT will be produced for:

- TEAEs
- TEAEs by worst NCI-CTCAE Grade (Any, ≥ 3 , ≥ 4 and 5)
- Study intervention related TEAEs by worst NCI-CTCAE Grade (Any, ≥ 3 , ≥ 4 and 5)
- Avelumab related TEAEs by worst NCI-CTCAE Grade (Any, ≥ 3 , ≥ 4 and 5)
- Combination drug related TEAEs by worst NCI-CTCAE Grade (Any, ≥ 3 , ≥ 4 and 5)
- TEAEs leading to death
- Study drug related TEAEs leading to death (PA only)
- Avelumab related TEAEs leading to death
- Combination drug related TEAEs leading to death

- For PA only, TEAEs excluding SAEs, with frequency $\geq 5\%$ in any study intervention arm. “Number of subjects with at least one event” represents the number of participants with at least one AE among AEs which frequency is $\geq 5\%$ in at least one study intervention arm. (clinicaltrials.gov and EudraCT requirement)

A TEAE table by SOC and PT displaying in separate columns the All TEAEs / Related TEAEs / Grade ≥ 3 TEAEs / Related Grade ≥ 3 TEAEs will be produced. Relationship refers to any study drug received.

Participant listings of AE details collected on the “Adverse Events Details” eCRF page will also be provided:

- AEs
- TEAEs with NCI-CTCAE Grade ≥ 3 and serious TEAEs

In all the AE listings, a flag will be added to indicate whether the AE or part of it is treatment-emergent or not. The AESIs will also be flagged.

For the DMC open session, the following tables will be produced:

- Overall summary of TEAEs and AESIs
- TEAEs by SOC, PT and relationship (any relationship and related to any study drug)
- TEAEs by worst grade, SOC, PT and relationship (any relationship and related to any study drug)

The tables above will be sorted in descending frequency of SOC and PT of the “any relationship” column and “Any Grade” when present.

15.1.2 Adverse Events Leading to Discontinuation of Study Intervention and Other Actions

The frequency (number and percentage) of participants with each of the following will be presented in a summary table:

- TEAEs leading to temporary discontinuation of study intervention (at least one study drug, avelumab, the combination drug, and both avelumab and the combination drug)
- Study intervention related TEAEs leading to temporary discontinuation of study intervention (at least one study drug, avelumab, the combination drug, and both avelumab and the combination drug)
- TEAEs leading to permanent discontinuation of study intervention (at least one study drug, avelumab, the combination drug, and both avelumab and the combination drug)
- Study intervention related TEAEs leading to permanent discontinuation of study intervention (at least one study drug, avelumab, the combination drug, and both avelumab and the combination drug)

- TEAEs leading to infusion rate reduction of study intervention (at least one study drug, avelumab, the combination drug, and both avelumab and the combination drug)
- Study intervention related TEAEs leading to infusion rate reduction of study intervention (at least one study drug, avelumab, the combination drug, and both avelumab and the combination drug)
- TEAEs leading to dose reduction of SG
- SG related TEAEs leading to dose reduction of SG

In addition, tables summarizing the frequency (number and percentage) of participants with TEAEs presented by SOC and PT will be prepared for:

- TEAEs leading to the permanent discontinuation of at least one study drug
- TEAEs leading to the permanent discontinuation of avelumab
- TEAEs leading to the permanent discontinuation of the combination drug
- Any study intervention related TEAEs leading to the permanent discontinuation of at least one study drug (PA only)
- Any study intervention related TEAEs leading to the permanent discontinuation of avelumab
- Any study intervention related TEAEs leading to the permanent discontinuation of the combination drug
- TEAEs leading to dose reduction of SG (PA only)
- SG related TEAEs leading to dose reduction of SG (PA only)
- TEAEs leading to infusion rate reduction (PA only)
- Study intervention related TEAEs leading to infusion rate reduction of at least one study drug (PA only)
- Study intervention related TEAEs leading to infusion rate reduction of avelumab (PA only)
- Study intervention related TEAEs leading to infusion rate reduction of combination drug (PA only)

A participant listing presenting all the TEAEs that lead to permanent or temporary discontinuation, reduction of dose or infusion rate will be provided.

Actions that can be taken by the Investigator to address an Adverse Event are collected by sections “Action(s) taken with Avelumab”, “Action(s) taken with Sacituzumab Govitecan”, “Action(s) taken with M6223”, “Action(s) taken with NKTR-255” on the “Adverse Events Details” eCRF page. It is possible that a TEAE leads to more than one action: for example, a drug interruption followed by drug withdrawal. In this case, the TEAE will be counted among the TEAEs leading to temporary discontinuation as well as among the TEAEs leading to permanent discontinuation.

15.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

15.2.1 Deaths

All deaths, deaths within 30 days after last dose of study intervention, death within 60 days after first dose as well as reason for death, will be tabulated based on information from the “Death” CRF page:

- Number of deaths (overall and by primary reason)
- Number of deaths within 30 days after last dose of study intervention (overall and by primary reason)
- Number of deaths within 60 days after first dose of study intervention (overall and by primary reason)

The primary reason of death can be:

- Progressive disease and/or disease related condition
- Event unrelated to study intervention
- Event related to avelumab only
- Event related the combination drug only (i.e., SG, M6223, NKTR-255 as applicable)
- Event related to avelumab and the combination drug
- Unknown
- Other reason
- Multiple reasons

In addition, date and cause of death will be provided in individual participant data listing together with selected dosing information (date of first / last administration and cumulative actual dose for avelumab and, if applicable, for the combination drug).

This listing will also include:

- AEs with fatal outcome (list preferred terms of AEs with outcome=fatal)
- Flag for death within 30 days of last study intervention
- Flag for death within 60 days of first study intervention

15.2.2 Serious Adverse Events

Tables summarizing the frequency (number and percentage) of participants with TEAEs by SOC and PT will be produced for:

- Serious TEAEs

- Study intervention related serious TEAEs (for PA only)
- Avelumab related serious TEAEs
- Combination drug related serious TEAEs

In addition, as indicated in Section 15.1.1, all serious TEAEs will be listed together with the TEAEs with grade ≥ 3 .

Serious AEs will be identified by looking at response to the section “Serious adverse event” in the “Adverse Events Details” eCRF page. A flag will be added to indicate whether the TEAE is an AESI (see Section below).

For the DMC open session, a table summarizing the frequency of participants with serious TEAEs by SOC, PT and relationship (any relationship and related to any study drug) will be produced. Sort by descending frequency of SOC and PT of the “any relationship” column.

15.2.3 Other Significant Adverse Events

The following AESI were defined in the protocol:

- Infusion-related reactions
- Immune-related adverse events
- Thromboembolic events
- Cytokine release syndrome
- QT interval prolongation

Pre-specified search lists of MedDRA Preferred Terms to be used for the identification of the AESI listed above will either be provided by the Sponsor or obtained by selecting specific SMQs directly from the MedDRA dictionary. These lists of PTs may need to be updated as new MedDRA dictionary versions become available during the study.

15.2.3.1 Infusion-related Reactions

IRRs are defined as AEs with preferred terms on a pre-specified list of MedDRA preferred terms and divided into two subgroups: reactions, signs and symptoms.

Reactions of IRR: include the PTs drug-hypersensitivity, hypersensitivity, Type-1-hypersensitivity and anaphylactic reaction. These PTs should be considered when the AE onset is on the day of any study drug infusion (during or after) or within one day thereafter.

Signs and symptoms of IRRs: include the PTs pyrexia, chills, flushing, hypotension, dyspnea, wheezing, back pain, abdominal pain and urticaria. These PTs should be considered if the AE onset occurs on the day of any study drug infusion (during or after) and resolves (i.e., the AE outcome is “Recovered/resolved” or “Recovered/resolved with sequelae”) within 2 days thereafter.

Data related to the IRR will be collected in the “Adverse Event Details” eCRF page whereas the details related to the reactions, signs and symptoms as part of the overall IRR will be provided in the “Infusion Related Reaction” eCRF form.

The overall summary of IRR table, based on data collected on the “Adverse Event Details” eCRF page, will include the following summaries:

- IRR
- Serious IRR
- IRR with NCI-CTCAE Grade ≥ 3 , = 3, = 4, = 5
- IRR leading to permanent discontinuation of avelumab
- IRR leading to permanent discontinuation of the combination drug
- IRR leading to permanent discontinuation of avelumab and the combination drug
- IRR leading to death

A frequency table of IRR AEs by worst grade, and PT will also be provided (overall and by IRR subgroup). This table will be based on data collected on the “Infusion Related Reaction” eCRF page.

For PA, IRR AEs will be flagged on the AE listings based on “Adverse Event Details” eCRF form. An IRR AE listing will also be produced with details based on “Infusion Related Reaction” eCRF form.

For the IA, for JBM subjects only, the serious IRRs will be listed together with the information related to the premedication (premedication preferred term, timing of administration) and to the treatment infusion (timing of avelumab and/or combination drug administration, as applicable).

15.2.3.2 Immune-related Adverse Events

irAEs will be identified programmatically. AEs which satisfy all of the following criteria will be flagged as immune-related:

- 1- The AE PT matches a PT on the list of pre-selected MedDRA terms
- 2- The AE onset occurs after the first administration of any study drug and no more than 90 days after the last dose of any study drug
- 3- On the “Adverse Events Details” eCRF page, the question “Were Corticosteroids, Immunosuppressants, or Hormonal Therapy (e.g., Thyroid Hormone, Insulin) Applied?” has the answer “Yes” selected
- 4- On the “irAE Specific Questions” eCRF page, either:

- a) The question “Does any of the following provide a clear etiology for the event, other than irAE?” has the answer “No” selected, indicating that the AE is not attributable to underlying cancer disease/PD, prior or concomitant medications/procedures, nor another medical condition such as an infection or pre-existing disease

OR

- b) The question “Was a biopsy performed?” has the answer “Yes, is the histopathology/biopsy consistent with an immune-mediated event?” selected which is also answered “Yes”

In the case that criteria (1) through (3) are met, and entries for condition (4) are missing, the following rules will be applied:

- If the answer to “Does any of the following provide a clear etiology for the event, other than irAE?” (4a) is missing, the event will be considered as irAE (irrespective of biopsy results).
- If the answer to “Yes, is the histopathology/biopsy consistent with an immune-mediated event?” (4b) is missing, or if no biopsy was performed, and condition (4a) is not satisfied (i.e. “Yes” is selected as the answer to the question “Does any of the following provide a clear etiology for the event, other than irAE?”), the event will be considered as a non-irAE.

PTs will be compiled into categories: immune-related rash, immune-related colitis, immune-related pneumonitis, immune-related hepatitis, immune-related nephritis and renal dysfunction, Immune-related endocrinopathies (adrenal insufficiency, hypogonadism, pituitary dysfunction, Type 1 diabetes mellitus, thyroid disorders), and other immune-related adverse events (e.g., immune-related pancreatitis, immune-related myocarditis, myositis, Guillain-Barré syndrome, uveitis, myasthenia gravis/myasthenic syndrome). The listed categories may need to be updated based on medical evaluation during the study.

For PA only, the overall summary of irAEs table will include the frequency (number and percentage) of participants with each of the following:

- irAE
- Serious irAE
- irAE with NCI-CTCAE Grade ≥ 3 , = 3, = 4, = 5
- irAE leading to permanent discontinuation of avelumab
- irAE leading to permanent discontinuation of the combination drug
- irAE leading to permanent discontinuation of avelumab and the combination drug
- irAE leading to death

A frequency table of irAEs by worst NCI-CTCAE grade, category, and PT will be provided.

Immune-related AEs will also be listed for PA.

15.2.3.3 Thromboembolic Events

The Standardized MedDRA Query (SMQ) “Embolic and thrombotic events” will be used to identify the thromboembolic events.

The number and percentage of participants who experience thromboembolic events with onset during the on-treatment period (see Section 9.7) will be summarized by worst NCI-CTCAE grade, SOC and PT.

Thromboembolic AEs will also be listed for PA.

15.2.3.4 Cytokine Release Syndrome

Any “Cytokine Release Syndrome” or “Cytokine Storm” event reported after the adoption of protocol version 3 will have the American Society for Transplantation and Cellular Therapy (ASTCT) grade according to Lee (2018) collected on the “Adverse Event Details” eCRF page. In addition, the “Cytokine Release Syndrome” eCRF page will collect fever, hypoxia, and hypotension symptom details. The number and percentage of participants who experienced cytokine release syndrome events will be summarized by worst ASTCT grade and PT. For those “Cytokine Release Syndrome” or “Cytokine Storm” events that occurred before protocol version 3 was adopted, they will be reported too but the ASTCT grade will be missing as well as the additional details collected on the “Cytokine Release Syndrome” eCRF page.

For PA, cytokine release syndrome events will be flagged on the AE listings based on “Adverse Event Details” eCRF page. A dedicated listing will also be produced with details based on “Cytokine Release Syndrome” eCRF page.

15.2.3.5 QT Interval Prolongation

The Standardized MedDRA Query (SMQ) “Torsade de pointes/QT prolongation” will be used to identify the AEs indicating QT Interval Prolongation.

The number and percentage of participants who experience the QT prolongation events with onset during the on-treatment period (see Section 9.7) will be summarized by worst NCI-CTCAE grade, SOC and PT.

QT interval prolongation AEs will also be listed for PA.

15.3 Clinical Laboratory Evaluation

Laboratory assessments such as hematology and coagulation, biochemistry, urinalysis as well as other tests such as pregnancy tests, tests for thyroid function, hepatitis B and C, etc., will be performed by local laboratories.

Laboratory results will be classified according to the NCI-CTCAE criteria version 5.0. Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: below normal limit, within normal limits and above normal limit (according to the laboratory normal ranges).

Baseline and change from baseline definitions are in Section 9.2.

Values below the detection limit will be imputed by half of the detection limit.

In case just a text value with an "> x" is reported it will be analyzed as +1 significant digit, e.g., "> 7.2 mmol" will be analyzed as 7.3.

All laboratory values will be reported in SI units.

For **calcium**, CTCAE grading is based on corrected calcium and ionized calcium, if available. Corrected calcium is calculated from albumin and total calcium as follows

Corrected calcium (mmol/L)
$$= \text{measured total calcium (mmol/L)} + 0.02(40 - \text{serum albumin [g/L]}).$$

Creatinine clearance is not collected among the biochemistry parameters but, as it will be needed to support the by subgroup efficacy analyses, the value at baseline will be calculated by the Cockcroft-Gault formula as

$$\text{Creatinine clearance (mL/min)} = \frac{\{[140 - \text{age}(\text{year})] \times \text{weight}(\text{kg})\}}{72 \times \text{serum creatinine (mg/dL)}} \times [0.85 \text{ for females}].$$

For the calculation of creatinine clearance:

$$\text{Serum creatinine (mg/dL)} = \text{serum creatinine (umol/L)} / 88.42.$$

Age at the date of the creatinine collection (only the integer part should be considered) and the closest measurement of weight to the date of creatinine collection will be used 9.5.

Parameters with NCI-CTC grades available:

Abnormalities classified according to NCI-CTCAE toxicity grading version 5.0 will be described using the worst on-treatment grade. Measurements from both scheduled and unscheduled visits performed during the on-treatment period (see Section 9.7) will be considered for the worst on-treatment grade derivation.

For those parameters which are graded with two directions of toxicities such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (e.g., hypokalemia) grades at baseline and post baseline will be set to 0 when the variables are derived for summarizing high direction toxicity (e.g., hyperkalemia), and vice versa.

The denominator to calculate percentages for each laboratory parameter will be the number of participants in each analyzed group. Participants without on-treatment results for a given parameter will be presented in the “Missing” category and will contribute to the denominator. Grade 0 is not defined per NCI-CTCAE but will be used in derivations for simplicity to indicate that evaluable measurements are available.

For PA only, the following summary tables will be produced:

- The summary of laboratory parameters by worst NCI-CTCAE grade table will include number and percentage of participants with Grade ≥ 0 , ≥ 1 , ≥ 3 , 4 laboratory abnormalities during the on-treatment period. Participants with no abnormality will be displayed under Grade 0.
- The shift table will summarize baseline NCI-CTCAE grade versus the worst on-treatment CTCAE grade.
- In case of gradings involving baseline measurements for the identification of grades during the on-treatment period, another shift table will present baseline normal and abnormal. Normal will include measurements below and within normal range (direction increase), or measurements within and above normal range (direction decrease). In case of missing baseline values, the on-treatment grades will be generated assuming the baseline was normal. For the eosinophils parameter, when baseline is missing, on-treatment values are graded considering only the upper limit of normal.

The above analyses apply to hematology, core biochemistry evaluations and prothrombin international normalized ratio (INR) which can be graded per CTCAE.

Parameters with NCI-CTC grades not available:

Parameters which cannot be graded per CTCAE criteria will be summarized as frequency (number and percentage) of participants by

- worst on-treatment value
- shift from baseline to worst on-treatment value (for PA only)

The worst on-treatment value is going to be derived considering the reference range categories (i.e., low, normal, high). Measurements from both scheduled and unscheduled visits performed during the on-treatment period (see Section 9.7) will be considered for the worst on-treatment value derivation.

The denominator to calculate percentages for each laboratory parameter is the number of participants in each analyzed group. Participants without on-treatment results for a given parameter will be presented in the “Missing” category and will contribute to the denominator.

Following figures will be provided for each laboratory parameter:

- Boxplots of the laboratory values by study intervention group by timepoint
- Boxplots of the CFB of the laboratory values by study intervention group by timepoint

Boxplots for laboratory parameters where toxicity grades are defined based on the ratio of the parameter values and the upper limit of normal (ULN) will not be displayed using the unit of measurement but instead using the ratio of the measured value over ULN. This comprises ALT, AST, bilirubin, creatinine, creatine kinase. All graphical displays of the absolute values will be shown on a log-scale. In case there are less than 5 participants within a study intervention group who provided the values of a parameter for a specific visit, the boxplot for this group will not be presented. Baseline as well as on-treatment values from scheduled visits will be presented. For JB100, only those visits with a correspondence to JBM visits will be presented.

In [Appendix 4](#) all the hematology, core biochemistry, coagulation and hormonal parameters are listed together with the direction of abnormality. All the parameters listed are to be presented in tables whereas those to be presented in the boxplots are flagged. Additional parameters among those not currently selected for the boxplots may be presented as well in case they become of clinical interest during the study.

Liver function tests:

ALT, AST, and total bilirubin are used to assess possible drug induced liver toxicity. The ratios of test result over ULN will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of participants with each of the following during the on-treatment period will be summarized by study intervention group for PA:

- $ALT \geq 3 \times ULN$, $ALT \geq 5 \times ULN$, $ALT \geq 10 \times ULN$, $ALT \geq 20 \times ULN$
- $AST \geq 3 \times ULN$, $AST \geq 5 \times ULN$, $AST \geq 10 \times ULN$, $AST \geq 20 \times ULN$
- $(ALT \text{ or } AST) \geq 3 \times ULN$, $(ALT \text{ or } AST) \geq 5 \times ULN$, $(ALT \text{ or } AST) \geq 10 \times ULN$, $(ALT \text{ or } AST) \geq 20 \times ULN$
- $TBILI \geq 2 \times ULN$
- Concurrent $ALT \geq 3 \times ULN$ and $TBILI \geq 2 \times ULN$
- Concurrent $AST \geq 3 \times ULN$ and $TBILI \geq 2 \times ULN$
- Concurrent $(ALT \text{ or } AST) \geq 3 \times ULN$ and $TBILI \geq 2 \times ULN$
- Concurrent $(ALT \text{ or } AST) \geq 3 \times ULN$ and $TBILI \geq 2 \times ULN$ and $ALP > 2 \times ULN$
- Concurrent $(ALT \text{ or } AST) \geq 3 \times ULN$ and $TBILI \geq 2 \times ULN$ and $ALP \leq 2 \times ULN$ or missing

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, i.e., a participant with an elevation of AST $\geq 10 \times \text{ULN}$ will also appear in the categories $\geq 5 \times \text{ULN}$ and $\geq 3 \times \text{ULN}$.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created, with different symbols for different study intervention groups, by graphically displaying a figure with log-scale transformed axes presented as

- peak serum ALT(/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT = $3 \times \text{ULN}$ and total bilirubin = $2 \times \text{ULN}$.

In the eDISH plot, there will be three pages presenting the eDISH plot for Group A and B, Group A and C, and Group A and D respectively. If the data from JB100 will be used, those will be presented as well as a third group within each page.

In addition, for PA, the following by-participant lab value listings presenting all corresponding collected information on the eCRF will be provided:

- Listing of all hematology values
- Listing of all biochemistry values
- Listing of participants with any AST or ALT $\geq 3 \times \text{ULN}$ or Total Bilirubin $\geq 2 \times \text{ULN}$
- Listing of all coagulation values
- Listing of all urinalysis values
- Listing of hormonal test results (based on "Hormonal Test 1 & 2" eCRF pages)
- Listing of Hepatitis B and Hepatitis C test results (based on "Serology" eCRF page, JB100 data will not be presented)
- Listing of pregnancy test results (based on "Pregnancy Test" eCRF page, JB100 data will not be presented)

15.4 Vital Signs

Analysis: PA

Vital signs data parameters as collected in the "Vital Signs eCRF page" during on-treatment period (on-treatment period defined in Section 9.7) will be summarized by study intervention group.

Shift tables of the maximum changes in vital signs measurements from baseline will be provided, vital signs measurements will be grouped as follows:

Table 6 Categories of Maximum Change from Baseline in Vital Signs

Body temperature increase from baseline < 37 °C, 37 - < 38°C, 38 - < 39°C, 39 - < 40°C, ≥ 40°C	< 1°C, 1 - < 2°C, 2 - < 3°C, ≥ 3 °C
---	-------------------------------------

Heart rate increase from baseline < 100 bpm; ≥ 100 bpm	≤ 20 bpm, > 20 - ≤ 40 bpm, > 40 bpm
Heart rate decrease from baseline < 100 bpm; ≥ 100 bpm	≤ 20 bpm, > 20 - ≤ 40 bpm, >40 bpm
SBP increase from baseline < 140 mmHg; ≥ 140 mmHg	≤ 20 mmHg, > 20 - ≤ 40 mmHg, > 40 mmHg
SBP decrease from baseline < 140 mmHg; ≥ 140 mmHg	≤ 20 mmHg, > 20 - ≤ 40 mmHg, > 40 mmHg
DBP increase from baseline < 90 mmHg; ≥ 90 mmHg	≤ 20 mmHg, > 20 - ≤ 40 mmHg, > 40 mmHg
DBP decrease from baseline < 90 mmHg; ≥ 90 mmHg	≤ 20 mmHg, > 20 - ≤ 40 mmHg, > 40 mmHg
Respiration rate increase from baseline < 20 breaths/min; ≥ 20 breaths/min	≤ 5 breaths/min, > 5 - ≤ 10 breaths/min, > 10 breaths/min
Respiration rate decrease from baseline < 20 breaths/min; ≥ 20 breaths/min	≤ 5 breaths/min, > 5 - ≤ 10 breaths/min, >10 breaths/min

All vital signs, including weight, will also be listed, baseline values and post-baseline values collected during the on-treatment period will be flagged in the listing.

15.5 Other Safety or Tolerability Evaluations

15.5.1 ECG

Single 12-lead ECG will be obtained using an ECG machine that automatically calculates the heart rate (HR), PR, RR, QRS, QT, QT corrected intervals (Fridericia's, Bazett's or other correction method). ECG results will be read locally and only one reading per visit will be performed. ECG parameter values will be collected in the "Electrocardiogram" eCRF page.

The analysis of QT data is complicated by the fact that the QT interval is highly correlated with heart rate. Because of this correlation, formulas are routinely used to obtain a corrected value, denoted QTc, which is independent of heart rate. This QTc interval is intended to represent the QT interval at a standardized heart rate. Several correction formulas have been proposed in the literature. For this analysis the QT interval corrected for heart rate by the Fridericia's formula, QTcF (msec), will be used. It is defined as

$$QTcF = \frac{QT}{\sqrt[3]{RR}}$$

where RR represents the RR interval of the ECG, in seconds, and can be estimated as 60/Heart Rate (beats/min) in case the RR interval (msec) was not collected in the eCRF.

In case the QTcF value cannot be derived using the formula above, the value provided in the eCRF will be used, if available.

Descriptive statistics (absolute and CFB) over time for 12-lead ECG parameters will be provided for treated subjects in Group D. Parameters to be included are: derived QTcF, HR, PR, RR, QT

and QRS duration. Baseline, Week 1 Day 1 Predose as well as on-treatment values from scheduled visits will be presented.

A listing presenting the data collected from ECG measurements will be produced. Values for the derived QTcF will also be included. Data from all the treated subjects will be presented as well as JB100 data, if suitable.

Additionally, the incidence of QT prolongation events as identified via SMQ will be evaluated as per Section 15.2.3.5.

16 Analyses of Other Endpoints/Estimands

16.1 Pharmacokinetics

PK evaluation will be performed by the Clinical PK/PD Group, Merck Healthcare KGaA, Darmstadt, Germany, or by a Contract Research Organization selected by the Sponsor.

16.1.1 Descriptive Statistics of PK Concentration Data

PK measurements will be descriptively summarized using: number of non-missing observations (n), arithmetic mean (Mean), standard deviation (StDev), coefficient of variation (CV%), minimum (Min), median (Median) and maximum (Max).

Descriptive statistics will only be calculated for $n > 2$ in which a measurement of BLQ represents a valid measurement and will be taken as zero for summary statistics of PK concentration data. In case $n \leq 2$, individual data will be presented (min, max) in summary tables.

Descriptive statistics of PK concentration data will be calculated using values with the same precision as the source data and rounded for reporting purposes only. PK concentrations will be carried over with full precision as provided in the source data without any rounding applied to CDISC SDTM PC and ADaM PC domains.

The following conventions will be applied when reporting descriptive statistics of PK concentration data:

n	0 decimal place
Mean, Min, Median, Max, StDev:	3 significant digits
CV%:	1 decimal place

16.1.2 Descriptive Statistics of PK Parameter Data

PK parameter data will be descriptively summarized using: n, Mean, StDev, CV%, Min, Median, Max, geometric mean (GeoMean), the geometric coefficient of variation (GeoCV%) and the 95% CI for the GeoMean (LCI 95% GM, UCI 95% GM).

Descriptive statistics will only be calculated for a PK parameter when $n > 2$. In case $n \leq 2$, individual data will be presented (min, max) in summary tables.

PK parameters read directly from the measurements (i.e., C_{trough} , C_{eoi}) will be reported with the same precision as the source data. Descriptive statistics of PK parameter data will be calculated using full precision on the derived parameter value and will be rounded for reporting purposes only.

PK parameters will be provided with full precision, without any rounding applied to CDISC SDTM PP and ADaM PP domains.

The following conventions will be applied when reporting descriptive statistics of PK parameter data:

n	0 decimal place
Mean, Min, Median, Max, GeoMean, 95% CI, StDev:	3 significant digits
CV%, GeoCV%:	1 decimal place

16.1.3 General Specifications for PK Concentration and PK Parameter Data

Predose samples that occur before the first drug administration will be assigned a time of 0 hours, as if the sample had been taken simultaneously with the study intervention administration. The same applies to the pre-dose sample for a multiple dose study.

Predose or trough samples which have been taken after the subsequent dosing will be treated as a protocol deviation. The resulting concentrations will be included in concentration listings but excluded from descriptive statistics of concentrations and from PK parameter estimation.

Values below the lower limit of quantification (BLQ) will be taken as zero for summary statistics of PK concentration data, PK parameter estimation (e.g. AUC) and for graphical presentations.

In case profiles have a measurable pre-dose concentration for the first dose which is less than or equal to 5% of the C_{eoi} value, the participant's data for that dose will be included in the descriptive and/or PK analyses without any adjustments. If the pre-dose value is greater than 5% of the C_{eoi} , the participant's data for that dose will be listed but excluded from descriptive and/or PK evaluation after agreement with the sponsor; data for subsequent doses will not be excluded for this reason.

Missing concentrations (e.g. no sample, insufficient sample volume for analysis, no result or result not valid) will be reported and used generally as "N.R.". A participant who withdraws prior to the last planned observation will be included in the analyses up to the time of discontinuation.

PK concentrations which are erroneous due to a sampling processing or analytical error (as documented in the bioanalytical report) may be excluded from the PK analysis if agreed by the Sponsor. In this case the rationale for exclusion will be provided in the CSR. Any other PK concentrations that appear implausible to the Clinical Pharmacologist/Clinical PK/PD Scientist will not be excluded from the analysis. Any implausible data will be documented in the CSR.

If important protocol deviations occurred likely to affect the PK profile of participants as described below, the impacted concentrations and PK parameters will be handled as described below.

Any PK concentrations or PK parameters excluded from summary statistics will be included in participants listings and flagged; a reason for exclusion will be detailed in the CSR (e.g. a footnote or a table of exclusions). Any flags should be included in the study specific CDISC data sets.

PK concentrations and PK parameters excluded from summary statistics will not be included in mean/median figures. Mean/median plots will only contain values where $n > 2$. In case $n \leq 2$, individual participant profiles may be included in mean/median plots.

In case of a dose change in the current dose and/or prior dose of investigational product, affected concentrations and parameters of the affected PK analyte will be flagged and included in listings, but will be excluded from statistical summaries. No data will be excluded for concomitant medication usage.

For samples collected outside the following time windows, concentrations will be flagged and included in the concentration summary and mean/median concentration plots:

- Predose samples that are collected more than 2 hours prior to the dose and/or outside of approximately $\pm 20\%$ of the scheduled dosing interval (τ) relative to the previous dose (defined as outside ± 3 calendar days for the 14-day τ interval for avelumab and M6223, ± 6 calendar days for the 28-day τ interval for NKTR-255, ± 1 calendar day for the 7-day τ interval for SG analytes, and ± 3 calendar days for the 14-day τ interval for SG analytes)
- EOI samples that are collected before EOI or later than approximately $+20\%$ of the actual infusion end time (defined as outside $+10$ min for NKTR-255 and $+20$ min for all other analytes)
- Postdose samples at 24 hours and 168 hours that are collected outside of approximately $\pm 20\%$ of their scheduled time point (defined as outside ± 5 hours for the 24-hour time point and ± 34 hours for the 168-hour time point)

In case there are duplicate PK samples for a scheduled time point, the results will be handled as follows: If the actual sample times for the duplicate samples are different, the concentration for the sample collected closest to the scheduled time point will be retained in the analysis, and the other sample(s) will be flagged and excluded. If the actual sample times for the duplicate samples are identical, the concentrations will be averaged and the mean concentration will be used in the analysis, provided that the percent difference (%Diff) of the concentrations C1 and C2 [estimated as $\%Diff = (C1 - C2) / \text{Mean}(C1, C2) * 100$] is less than or equal to $\pm 20\%$; otherwise the duplicate concentrations will be flagged and excluded.

16.1.4 Estimation of Pharmacokinetic Parameters

The computer program Phoenix® WinNonlin® Version 8.3, or higher (Certara, L.P., Princeton, New Jersey, USA) will be used to derive PK parameters.

Phoenix® WinNonlin® may be used to generate PK parameters, and/or produce tables, listings and figures. The statistical software SAS® (Statistical Analysis System, SAS-Institute, Cary, North Carolina, USA, windows Version 9.4 or higher) may be used to generate additional PK parameters, produce tables, listings and figures.

16.1.5 Estimation of Pharmacokinetic Parameters in Serum or Plasma

PK parameters will be calculated using the actual elapsed time since dosing. In cases where the actual sampling time is missing, calculations may be performed using the nominal time. Details (e.g. number of samples, participants affected) will be described in the CSR. In case actual dosing time is missing, scheduled time might be used for noncompartmental (NCA) analysis after performance of adequate plausibility checks and agreement with the Sponsor. Decision and rationale should be included in the CSR. Otherwise, there will be no further imputation of missing data.

If less than 3 quantifiable post-dose concentrations are available within a PK profile and/or the EOI sample and/or the predose sample (on a multiple dose day) are missing, the profile will be excluded from PK analysis or only limited PK parameters will be calculated.

The serum/plasma PK parameters specified in Table 7 will be calculated, if possible, for avelumab, M6223, NKTR-255, and the SG analytes (free SN-38, total SN-38, and total antibody), as appropriate:

Table 7 Pharmacokinetic Parameters to be Calculated

Symbol	Definition	Calculation
C _{ei}	The observed concentration at the end of the infusion period.	Obtained directly from the concentration-time profile
C _{trough}	The concentration observed at the end of a dosing interval immediately before next dosing.	Obtained directly from the concentration-time profile
AUC _{0-tlast}	The AUC from time zero (= dosing time) to the time of the last quantifiable concentration (t _{last}).	Calculated using the mixed log-linear trapezoidal rule (linear up, log down). Calculated for avelumab and M6223 on C1D1 and C2D1 in Group C and for avelumab and NKTR-255 on C1D1 and C2D1 in Group D when possible.
AUC _τ	The AUC over the dosing interval τ (i.e. 336 hours for avelumab and M6223, and 672 hours for NKTR-255, when possible).	Calculated using the mixed log-linear trapezoidal rule (linear up, log down). For single dose, AUC _τ is calculated as a partial area with the defined time range. In multiple dose profiles, AUC _τ is calculated from a pre-dose time point to the end of the dosing interval. Calculated for avelumab and M6223 on C1D1 and C2D1 in Group C and for

		avelumab on C1D1 and C2D1 in Group D when possible.
λ_z	Terminal first order (elimination) rate constant.	Determined from the terminal slope of the log-transformed concentration curve using linear regression on terminal data points of the curve. The EOI sample may be included in the estimation of this parameter, if warranted. Calculated for avelumab and M6223 on C1D1 and C2D1 in Group C and for avelumab on C1D1 and C2D1 in Group D when possible.
$t_{1/2}$	The terminal half-life.	$t_{1/2} = \ln(2) / \lambda_z$, where λ_z is the terminal first order (elimination) rate constant determined from the terminal slope of the log-transformed concentration curve using linear regression on terminal data points of the curve. Calculated for avelumab and M6223 on C1D1 and C2D1 in Group C and for avelumab on C1D1 and C2D1 in Group D when possible.
$AUC_{0-\infty}$	The AUC from time zero (= dosing time) up to infinity with extrapolation of the terminal phase.	$AUC_{0-\infty} = AUC_{0-t_{last}} + C_{last\ pred} / \lambda_z$, where $C_{last\ pred}$ is the predicted concentration at the t_{last} , calculated from the log-linear regression line for λ_z determination. Calculated for avelumab and M6223 on C1D1 in Group C and for avelumab on C1D1 in Group D when possible.
CL	The total body clearance following intravenous administration.	Calculated as $CL = \text{Dose} / AUC_{0-\infty}$ for single dose, and as $CL = \text{Dose} / AUC\tau$ for multiple dose. Calculated for avelumab and M6223 on C1D1 in Group C, and for avelumab on C1D1 and C2D1 on C1D1 in Group D when possible.
V_z	The volume of distribution during the terminal phase following intravenous administration.	Calculated as $V_z = \text{Dose} / (AUC_{0-\infty} \times \lambda_z)$ for single dose, and as $V_z = \text{Dose} / (AUC\tau \times \lambda_z)$ for multiple dose. Calculated for avelumab and M6223 on C1D1 in Group C and for avelumab on

		C1D1 and C2D1 on C1D1 in Group D when possible.
V_{ss}	The apparent volume of distribution at steady state following intravenous administration	Calculated as $V_{ss} = MRT_{0-\infty} \times CL$ for single dose, and as $V_{ss} = MRT_{ss} \times CL$ for multiple dose. Calculated for avelumab and M6223 on C1D1 in Group C and for avelumab on C1D1 and C2D1 on C1D1 in Group D when possible.

Additional PK parameters may be calculated as appropriate.

For calculation of dose-related avelumab, M6223, and NKTR-255 PK parameters, no dose adjustments (for salt forms, etc.) are needed.

Units for PK parameter output will be based on concentration units used in the study, unless otherwise specified. For the SG analyte, total antibody, the units for PK parameters will be based on the concentration unit ug/mL, and for NKTR-255, it will be based on ng/mL. In case concentration data units are changed, PK parameters will be reported using consistent units throughout study outputs.

The parameters C_{eoi} , and C_{trough} will be obtained directly from the concentration-time profiles. C_{trough} will be listed and flagged but excluded from the PK parameter summaries if it is collected outside of approximately $\pm 20\%$ of the scheduled τ interval relative to the previous dose (defined as outside ± 3 calendar days for the 14-day τ interval for avelumab and M6223, ± 6 calendar days for the 28-day τ interval for NKTR-255, ± 1 calendar day for the 7-day τ interval for SG analytes, and ± 3 calendar days for the 14-day τ interval for SG analytes). For C_{eoi} , if it is collected before EOI or later than approximately $+20\%$ of the actual infusion end time (defined as outside $+10$ min for NKTR-255 and $+20$ min for all other analytes), it will be listed and flagged but excluded from the PK parameter summaries.

In cases where the actual observation time at τ is not equal to the scheduled observation time, AUC_{τ} will be calculated based on the estimated concentration at scheduled τ , and not the concentration at the actual observation time. However, in case interpolation to scheduled τ is not possible, and a suitable regression for extrapolation cannot be performed (see criterion for Rsq_{adj} below), partial areas will be calculated using the actual sampling time at scheduled τ , provided it is within approximately 10% of the scheduled sampling time (defined as ± 34 hours for $\tau = 336$ hours). If the actual sampling time is outside this interval, AUC_{τ} and related parameters based on the estimated concentration at scheduled τ will be listed, and flagged but excluded from the PK parameter summaries.

The following further rules might be applied for calculation of AUC_{τ} , after agreement with the Sponsor:

- In cases where BLQ concentrations occur at the end of the collection interval, these concentrations might be set to missing for calculations of partial AUCs. This is an exemption to the general BLQ handling rule and should be applied in case its application would result in estimation of implausible partial AUC values. Implausibility is considered in cases where partial AUCs are greater than $AUC_{0-\infty}$.
- If the pre-dose concentration within a steady-state concentration-time profile is missing, but a sample is available at tau that was collected within approximately 10% of the scheduled sampling time (± 34 hours for $\tau = 336$ hours), the concentration at tau will be substituted for the missing pre-dose concentration for calculation of PK parameters. The substituted concentration will be flagged and excluded from descriptive summaries of concentrations.

The following PK parameters will be calculated for avelumab, M6223, and NKTR-255 for diagnostic purposes and listed when available, but will not be summarized:

- First (λ_z low) and last (λ_z up) time point of the time interval of the log-linear regression to determine λ_z .
- Number of data points (N_λ) included in the log-linear regression analysis to determine λ_z .
- Goodness of fit statistic (Rsq_adj) for calculation of λ_z .
- AUC from time t_{last} extrapolated to infinity given as percentage of $AUC_{0-\infty}$ ($AUC_{extra\%}$).
- Span of interval over which $t_{1/2}$ was estimated/ $t_{1/2}$.

The regression analysis should contain data from at least 3 different time points in the terminal phase consistent with the assessment of a straight line on the log-transformed scale. Phoenix WinNonlin “best fit” methodology will be used as standard. If warranted, further adjustment may be made by the pharmacokineticist, after agreement with the Sponsor. The last quantifiable concentration greater than the lower limit of quantification (LLOQ) should always be included in the regression analysis.

If the coefficient of correlation (Rsq_adj) of λ_z is <0.800 , the λ_z and all related derived parameters (e.g. $t_{1/2}$, $AUC_{0-\infty}$, CL, etc.) will be listed, flagged, and included in the parameter outputs and excluded from descriptive statistics and further statistical evaluation. If $AUC_{extra\%} > 20.0\%$, $AUC_{0-\infty}$ and all related parameters will be listed, flagged, and included in the parameter outputs and excluded from descriptive statistics and further statistical evaluation. If the observation period over which the regression line is estimated (Span) is less than 2-fold, the resulting parameters will be listed, flagged, and included in the parameter outputs, and included in descriptive statistics and further statistical evaluation (unless other exclusion criteria apply).

16.1.6 Presentation of PK Concentration and PK Parameter Data

16.1.6.1 Listings and Tables

The following PK tables will be produced (PKAS):

- Descriptive statistics of concentrations by analyte, week, and measurement time, separately for each group
- Descriptive statistics of PK parameters by analyte and week for Group C and Group D (C1D1 and C2D1 doses only)
- Descriptive statistics of C_{trough} and C_{coi} by analyte and week, separately for each group

The following PK Listings will be produced (Safety Analysis Set):

- PK sample collection dates and times and analyte concentrations
- Individual concentrations by analyte, week, and measurement time, separately for each group
- Individual PK parameters by analyte and week for Group C and for Group D (C1D1 and C2D1 doses only)
- Individual C_{trough} and C_{coi} by analyte and week, separately for each group
- Concentrations and PK parameters excluded from the PKAS and reason for exclusion, for each group, analyte, and subject
- Phoenix WinNonlin Core Output

16.1.6.2 Graphical Summaries and Individual Plots

The following plots will be produced (PKAS):

- Overlaid individual concentration versus time plots; linear and semi-logarithmic; by analyte and day, for Group C and for Group D
- Arithmetic mean concentration time plots; linear ($\pm StDev$) and semi-logarithmic; using scheduled (nominal) time points by analyte and day, for Group C and for Group D
- Arithmetic mean $\pm StDev$ trough concentrations versus nominal time on a linear scale, by group and analyte.
- Boxplots of C_{trough} versus time point (ordinal scale) on a linear scale by group and analyte.
- Boxplots of C_{coi} versus time point (ordinal scale) on a linear scale by group and analyte.

16.2 Immunogenicity

16.2.1 Antidrug Antibody

Analysis: PA

Anti-drug antibody (ADA) sample results will be evaluated as follows:

1. avelumab ADA in Group A
2. avelumab ADA in Group B
3. avelumab ADA in Group C

4. avelumab ADA in Group D
5. SG ADA in Group B
6. M6223 ADA in Group C
7. NKTR-255 ADA in Group D

Samples collected after the on-treatment period (e.g., safety follow-up) will be included in the analyses. If the sample is confirmed positive for ADA, the titer will be measured. The ADA results will be derived based on the algorithm in the Table 8.

Table 8 **Algorithm for the Derivation of ADA Results**

Sample Screen Result	Confirmatory	Titer	ADA Result
Negative	NA	NA	Negative
NR	NA	NA	NR
Positive	Negative	NA	Negative
Positive	NR	NA	NR
Positive	Positive	Number	Number
Positive	Positive	TNR	Positive-TNR

NR = no result, NA = not applicable, TNR = titer no result.

Negative, number or positive-TNR are valid results while number and positive-TNR are considered as positive. Participants will be characterized into different categories based on the criteria in the table below (see Table 9).

Table 9 **Participants Characterized based on ADA Results**

Category	Definition	Participant at Risk (Denominator for Incidence)
Never positive	No positive results at any time point	Number of participants with at least one valid ADA result at any time point
Ever positive	At least one positive result at any time point	Number of participants with at least one valid ADA result at any time point
Pre-existing	A positive ADA result prior to study intervention	Number of participants with valid ADA baseline result
Treatment boosted	A positive ADA result prior to study intervention and the titer $\geq 8 \times$ baseline titer while on treatment	Number of participants with a valid ADA baseline result and at least one valid post-baseline ADA result
Treatment induced	Not positive prior to study intervention (includes missing or NR) and with at least one positive post-baseline result	Number of participants with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Treatment induced: Transient positive	If treatment induced participants have duration between first and last positive result < 16 weeks and without a positive evaluation within 16 weeks of the last ADA assessment.	Number of participants with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Treatment induced: Persistent positive	If treatment induced participants have duration between first and last positive result ≥ 16 weeks or have a positive evaluation within 16 weeks of the last ADA assessment	Number of participants with at least one valid post-baseline ADA result and without positive baseline ADA result (including missing, NR)
Treatment emergent	Treatment induced + Treatment boosted	Number of participants with at least one valid post-baseline result

The frequency and percentage of ADA categories will be tabulated separately for each study intervention category described at the beginning of this section.

A listing of ADA results in each study intervention category for all ADA ever-positive participants will be prepared.

16.2.2 Neutralizing Antidrug Antibody

Samples with a reportable ADA titer may also be tested in the neutralizing antibody (nAb) assays. nAb results are positive or negative in a single assay and only derived when not performed because ADA was negative (see Table 10). Participants will be characterized into different nAb categories based on the criteria in Table 11. Treatment boosted is not defined for nAb.

Table 10 **Algorithm for the Derivation of nAB Results**

ADA Confirmatory Result	nAb Result	Derived nAb Result
Negative	NA	Negative
NR	NA	NR
NA	NA	NA
Positive	NR	NR
Positive	Positive	Positive
Positive	Negative	Negative

ADA = antidrug antibody, NA = not applicable, nAb = neutralizing antibody, NR = no result.

Table 11 **Participants Characterized based on nAB Results**

Category	Definition	Subject at Risk (Denominator for Incidence)
Never positive	No nAb positive results at any time point	Number of participants with at least one valid ADA result at any time point
Ever positive	At least one nAb positive result at any time point	Number of participants with at least one valid ADA result at any time point
Pre-existing	A positive nAb result prior to treatment	Number of participants with valid ADA baseline result
Treatment emergent	Not nAb positive prior to treatment (includes missing or NR) and with at least one nAb positive post-baseline result	Number of participants with at least one valid ADA post-baseline result and without nAb positive baseline results (including missing, NR)
Transient positive	If treatment emergent participants have (a single nAb positive evaluation, or duration between first and last nAb positive result <16 weeks) and last ADA assessment not nAb positive.	Number of participants with at least one valid ADA post-baseline result and without nAb positive baseline results (including missing, NR)
Persistent positive	If treatment emergent participants have duration between first and last nAb positive result ≥16 weeks or a nAb positive evaluation at the last ADA assessment	Number of participants with at least one valid ADA post-baseline result and without nAb positive baseline results (including missing, NR)

ADA = antidrug antibody, nAb = neutralizing antibody, NR = no result.

The frequency and percentage of each nAb category will be tabulated separately for each point described at the beginning of Section 16.2.1. The only exemption will be for M6223 since there are no nAb assays for this drug.

A listing of nAb results in each treatment category for all ADA ever-positive participants (in either assay) will be provided.

16.2.3 Immunogenicity Subgroup Analysis of Efficacy

For the ADA ever-positive participants, a listing will be prepared containing, among other information, date and time of onset and last date and time of ADA positive results, as well as date and time of onset and last date and time of nAb positive results, BOR and BOR date, DoR, PFS

time or censoring time and reason for censoring, and OS time or censoring time and reason for censoring. Two listings will be produced, one considering the ADA results related to avelumab and a second one considering the ADA results related to the combination drugs.

For the ADA ever-positive participants, the percent CFB in target lesions as well as the first occurrence of a new lesion and participant off treatment will be displayed against time since start of treatment (months) in a spider plot. Additional symbols will indicate the first and last ADA positive result and, if applicable, the first and last nAb positive result. A plot will be produced for each study intervention category described at the beginning of Section 16.2.1.

16.2.4 Immunogenicity Subgroup Analysis of Adverse Events

A listing of immunogenicity data and TEAEs will be provided containing, among other information, the ADA status group, all dates with positive ADA results, all dates with positive nAb results, preferred term of TEAE, TEAE start and stop dates, CTCAE toxicity grade and flags for adverse event of special interest. Two listings will be produced, one considering the ADA results related to avelumab and a second one considering the ADA results related to the combination drugs.

Only participants who are pre-existing positive, transient treatment-emergent positive, or persistent treatment-emergent positive will be included in the listing.

16.2.5 Immunogenicity Subgroup Analysis of Drug Concentrations

A listing of immunogenicity data and drug concentrations will be provided containing, among other information, the ADA status group, all dates with positive ADA results, all dates with positive nAb results and the drug concentrations. Two listings will be produced, one considering the ADA results related to avelumab and a second one considering the ADA results related to the combination drugs.

Only participants who are pre-existing positive, transient treatment-emergent positive, or persistent treatment-emergent positive will be included in the listing.

In addition, for the treatment emergent participants, the C_{trough} over time will be presented. In case the C_{trough} value is below limit of quantification, zero will be presented instead. On the x-axis, the time relative to ADA seroconversion in days will be presented. Seroconversion is defined as the date where the first ADA positive result is observed or there is a titer booster (i.e., titer $\geq 8 \times$ baseline titer while on treatment). Plots will be created for avelumab ADA results for each study intervention group and for each combination drug ADA results. For SG, the analyte “hRS7 IgG and hRS7-SN38” will be analyzed as drug concentration.

17 References

- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res.* 2011 May;46(3):399-424. doi: 10.1080/00273171.2011.568786. Epub 2011 Jun 8. PMID: 21818162; PMCID: PMC3144483.
- Brookmeyer R, Crowley J. A Confidence Interval for the Median Survival Time. *Biometrics* 1982;38:29-41.
- Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*; 26, 404-413, 1934.
- Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45:228-247.
- Elizabeth A. Stuart. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statist. Sci.* 25 (1) 1 - 21, February 2010. <https://doi.org/10.1214/09-STS313>
- Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons 1980.
- Kenward MG, Roger JH, Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; 53: 983-997.
- Lee DW, Santomasso BD, Locke FL, Ghobadi A, Turtle CJ, Brudno JN, Maus MV, Park JH, Mead E, Pavletic S, Go WY, Eldjerou L, Gardner RA, Frey N, Curran KJ, Peggs K, Pasquini M, DiPersio JF, van den Brink MRM, Komanduri KV, Grupp SA, Neelapu SS. ASTCT Consensus Grading for Cytokine Release Syndrome and Neurologic Toxicity Associated with Immune Effector Cells. *Biol Blood Marrow Transplant.* 2019 Apr;25(4):625-638. doi: 10.1016/j.bbmt.2018.12.758. Epub 2018 Dec 25. PMID: 30592986.
- Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med.* 1987;317:1098. Walker A, Patrick A, Lauer M, Hornbrook M, Marin M, Platt R, Roger V, Stang P, Schneeweiss S. A tool for assessing the feasibility of comparative effectiveness research. *Comparative Effectiveness Research.* 2013;3:11-20 <https://doi.org/10.2147/CER.S40357>. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 *Bio-Algorithms and Med-Systems* 2013; 9(4): 183–9.

18 Appendices

Appendix 1 – SAP for DMC

See separate Appendix of this IAP.

Appendix 2 – JBM vs JB100 Eligibility Criteria

Category	Criterion n.	Medley Criteria	JB100 Inc./Exc. Code ^a	JB100 criteria	Comments/ Additional Information
Inclusion Criteria					
Age	1	Are ≥ 18 years of age at the time of signing the informed consent.	IN07A00	Age 18 years (≥ 20 years in Japan)	
Type of Participant and Disease Characteristics	2	Has histologically confirmed, unresectable locally advanced or metastatic urothelial carcinoma. Both transitional cell and mixed transitional/non- transitional cell histologies are allowed, but transitional cell carcinoma must be the predominant histology. documented Stage IIIA/IIIB with N1-N3, or Stage IV disease (per American Joint Committee on Cancer/International Union for Cancer Control Tumor Node Metastasis system, 8th edition) at the start of first line chemotherapy.	IN01A00 IN01A03	Histologically confirmed, unresectable locally advanced or metastatic transitional cell carcinoma of the urothelium. Documented Stage IV disease (per American Joint Committee on Cancer/International Union for Cancer Control Tumor Node Metastasis (TNM) system, 7th edition) at the start of first-line chemotherapy. Measurable disease prior to the start of first-line chemotherapy by RECIST v1.1.	
	3	Prior 1L chemotherapy must have consisted of at least 4 cycles and no more than 6 cycles of gemcitabine + cisplatin and/or gemcitabine + carboplatin. No other chemotherapy regimens are allowed in this study. (Note: Switch between platinum agents due for toxicity to complete a total of - 4 to 6 cycles first line platinum-based chemotherapy is acceptable).	IN02A00 IN02A03	Prior first-line chemotherapy must have consisted of at least 4 cycles and no more than 6 cycles of gemcitabine + cisplatin and/or gemcitabine + carboplatin. No other chemotherapy regimens are allowed in this study. The last dose of first-line chemotherapy must have been received no less than 4 weeks, and no more than 10 weeks, prior to randomization in the present study.	Inclusion 3 and 4 are combined in JB100. 4 was added while JB100 was ongoing. Since criterion 4 is less medically important than criterion 3, it will not be checked for participants who were evaluated using IN02A00.
	4	The last dose of first line chemotherapy must have been received no less than 4 weeks, and no more than 10 weeks, prior	IN02A03	See inclusion criterium 3	

	to randomization in the present study.			
5	Participants without progressive disease as per RECIST v1.1 guidelines (i.e., with an ongoing CR, PR, or SD) following completion of 4 to 6 cycles of 1L chemotherapy. Eligibility based on this criterion will be determined by Investigator review of pre chemotherapy and post chemotherapy radiological assessments (CT/MRI scans).	IN03A00 IN03A03	Patients without progressive disease as per RECIST v1.1 guidelines (i.e., with an ongoing CR, PR, or SD) following completion of 4 to 6 cycles of first-line chemotherapy. Eligibility based on this criterion will be determined by investigator review of pre-chemotherapy and post-chemotherapy radiological assessments (CT/MRI scans); see Study Overview, Section 3.1.	Criterion changed from independent review to investigator but it is not considered an issue.
6	All participants must provide archival formalin-fixed paraffin-embedded (FFPE) tumor tissues (ideally < 6 months old prior to participant enrollment) from most recent primary or metastatic tumor biopsy or resection obtained prior to treatment with first line chemotherapy but within 24 months prior to randomization, with no intervening systemic anticancer therapy between the time the tissue was obtained and initiation of first line chemotherapy. If an archival FFPE tissue block cannot be provided, 10 or more unstained slides will be acceptable. In addition, fresh baseline tumor samples (collected within 28 days before first dose) may be collected at Screening if archival tissue biopsy is not available. Provision of a tumor biospecimen is required for randomization into the study. The sample must be submitted to the Central Laboratory prior to randomization	IN04A00 IN04A03	Provision of a recent formalin-fixed, paraffin-embedded (FFPE) tumor tissue block (or subsection thereof) from the most recent primary or metastatic tumor biopsy or resection obtained prior to treatment with first line chemotherapy but within 24 months prior to randomization, with no intervening systemic anti-cancer therapy. If a FFPE tissue block cannot be provided, 15 freshly cut unstained slides (10 minimum) will be acceptable. If a suitable tissue sample is not otherwise available, then a de novo biopsy (core needle or excisional) must have been obtained for research purposes prior to randomization in this study. Note: tumor tissue from cytologic sampling (e.g., fine needle aspiration, including FFPE cell pellet material) or bone metastases are not acceptable and should not be submitted.	This criterion will not be considered as it is more to support CCI [REDACTED]

	and deemed acceptable to allow the patient to randomize. Refer to the Central Laboratory Manual for additional specifications.			
7	Participants who are willing and able to comply with scheduled visits, treatment plans, laboratory tests, and other study procedures.	IN06A00	Patients who are willing and able to comply with scheduled visits, treatment plans, laboratory tests, and other study procedures.	
8	Estimated life expectancy of at least 3 months.	IN08A00	Estimated life expectancy of at least 3 months.	
9	Eastern Cooperative Oncology Group (ECOG) performance status (PS) 0 or 1.	IN09A00	Eastern Cooperative Oncology Group (ECOG) performance status (PS) 0 or 1	
10	Adequate bone marrow function, including: Absolute neutrophil count (ANC) $\geq 1,500/\text{mm}^3$ or $\geq 1.5 \times 10^9/\text{L}$; Platelets $\geq 100,000/\text{mm}^3$ or $\geq 100 \times 10^9/\text{L}$; Hemoglobin $\geq 9 \text{ g/dL}$ (may have been transfused).	IN10A00	Adequate bone marrow function, including: Absolute neutrophil count (ANC) $\geq 1,500/\text{mm}^3$ or $\geq 1.5 \times 10^9/\text{L}$; Platelets $\geq 100,000/\text{mm}^3$ or $\geq 100 \times 10^9/\text{L}$; Hemoglobin $\geq 9 \text{ g/dL}$ (may have been transfused).	
11	Adequate renal function, defined as estimated creatinine clearance $\geq 30 \text{ mL/min}$ as calculated using the Cockcroft-Gault equation or by 24-hour urine collection for creatinine clearance or according to the local institutional standard method.	IN11A01 IN11A03	Adequate renal function, defined as estimated creatinine clearance $\geq 30 \text{ mL/min}$ as calculated using the Cockcroft-Gault equation (Appendix 4 of study protocol) or by 24-hour urine collection for creatinine clearance or according to the local institutional standard method.	Criterion was extended while JB100 study was ongoing.
12	Adequate liver function, including: Total serum bilirubin $\leq 1.5 \times$ upper limit of normal (ULN); Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) $\leq 2.5 \times$ ULN, or, for participants with documented metastatic disease to the liver, AST and ALT levels $\leq 5 \times$ ULN. Patients diagnosed with Gilbert's syndrome are eligible for enrollment.	IN12A00 IN12A03	Adequate liver function, including: Total serum bilirubin $\leq 1.5 \times$ upper limit of normal (ULN); Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) $\leq 2.5 \times$ ULN, or, for participants with documented metastatic disease to the liver, AST and ALT levels $\leq 5 \times$ ULN.	Criterion was extended while JB100 study was ongoing.

Sex and Contraception/Barrier Requirements	13	See protocol	IN13A00 IN14A00	See protocol amendments	
Informed Consent	14	Capable of giving signed informed consent, as indicated in Appendix 2 of the protocol, which includes compliance with the requirements and restrictions listed in the ICF and this protocol.	IN05A00	Evidence of a signed and dated informed consent document indicating that the patient (or a legally acceptable representative, as allowed by local guideline/practice) has been informed of all pertinent aspects of the study.	
History of HIV, HBV or HBC Note: We do not display these criteria because in JBM we would not have excluded them for sure and in JB100 there was no subject randomized who met these exclusion criteria.	15	Participants with known human immunodeficiency virus (HIV) infections are eligible if the following criteria are met: If clinically indicated participants must be stable on antiretroviral therapy (ART) for at least 4 weeks and agree to adhere to ART. If not clinically indicated, consult study Medical Monitor. Participants with HIV infection should have no evidence of documented multidrug resistance that would prevent effective ART. Have an HIV viral load of < 400 copies/mL at Screening. If prophylactic antimicrobial drugs are indicated, patient may still be considered eligible upon agreement with the study Medical Monitor.	EX21A02	Patients with positive test for human immunodeficiency virus (HIV) infection or known acquired immunodeficiency syndrome (AIDS) were excluded.	
	16	Participants with hepatitis B virus (HBV) and/or hepatitis C virus (HCV) infections are eligible if the following criteria are met: Patients with serologic evidence of chronic HBV infection must have an HBV viral load below the limit of quantification and be on a stable dose of antiviral therapy.	EX16A00 EX16A03	Patients with Hepatitis B virus (HBV) or hepatitis C virus (HCV) infection at screening (positive HBV surface antigen or HCV RNA if anti-HCV antibody screening test positive) were excluded.	The exclusion criterion changed during the JB100 study and became more restrictive so it can be used to check against the JBM criterion.

		Patients with a history of HCV infection should have completed curative antiviral treatment and require HCV viral load below the limit of quantification. Patients on concurrent HCV treatment should have HCV below the limit of quantification.			
Exclusion Criteria					
Medical Conditions	-	Not applicable	EX01A00	Patients whose disease progressed by RECIST v1.1 on or after first-line chemotherapy for urothelial cancer.	Criterion present only in JB100 but considered relevant for the JBM study.
	1	Participants with symptomatic central nervous system (CNS) metastases requiring steroids. Participants with diagnosed CNS metastases are eligible if they have completed treatment, recovered from the acute effects of radiation therapy or surgery prior to randomization and off steroid treatment 7 days before randomization, have discontinued corticosteroid treatment for these metastases for at least 4 weeks, and are neurologically stable.	EX05A00	Patients with known symptomatic central nervous system (CNS) metastases requiring steroids. Patients with previously diagnosed CNS metastases are eligible if they have completed their treatment and have recovered from the acute effects of radiation therapy or surgery prior to randomization, have discontinued corticosteroid treatment for these metastases for at least 4 weeks, and are neurologically stable.	
	2	Persisting toxicity related to prior therapy NCI CTCAE v5.0 Grade > 1; however, alopecia, sensory neuropathy Grade ≤ 2 is acceptable, or other Grade ≤ 2 adverse events not constituting a safety risk-based on the Investigator's judgment are acceptable.	EX06A00 EX06A03	Persisting toxicity related to prior therapy NCI CTCAE v4.0 Grade >1; however, alopecia, sensory neuropathy Grade ≤ 2 is acceptable, or other Grade ≤ 2 adverse events not constituting a safety risk based on the investigator's judgment are acceptable.	Criterion became less restrictive during the JB100 study. The last version corresponds to the criterion in the JBM study.
	3	Diagnosis of any other malignancy unless a complete remission without further recurrence was achieved and the participant was deemed to have been cured with	EX07A00 EX07A03	Diagnosis of any other malignancy within 5 years prior to randomization, except for adequately treated basal cell or squamous cell skin cancer, or	Criterion became less restrictive during the JB100 study but is still more restrictive than the one in JBM study.

		no additional therapy required or anticipated to be required. The exceptions to this criterion may include carcinoma in situ of cervical, colorectal, breast, or all localized prostate cancer that is not being treated. Possible exceptions to this criterion will be discussed on a by participant basis according to medical need.		carcinoma in situ of the breast or of the cervix, low-grade (Gleason ≤ 6) prostate cancer on surveillance without any plans for treatment intervention (e.g., surgery, radiation, or castration), or prostate cancer that has been adequately treated with prostatectomy or radiotherapy and currently with no evidence of disease or symptoms.	
	4	Active autoimmune disease that might deteriorate when receiving an immunostimulatory agent. Participants with diabetes type I, vitiligo, psoriasis, or hypo- or hyperthyroid disease not requiring immunosuppressive treatment are eligible.	EX09A00	Active autoimmune disease that might deteriorate when receiving an immunostimulatory agent. Participants with diabetes type I, vitiligo, psoriasis, or hypo- or hyperthyroid disease not requiring immunosuppressive treatment are eligible.	
	5	Current significant cardiac conduction abnormalities, including corrected QT interval (QTcF, corrected with Fridericia formula) prolongation of > 450 ms for male and >470 for female participants on triplicate 12 lead ECG or impaired cardiovascular function.			Criterion not to be checked
	6	Clinically significant (i.e., active) cardiovascular disease: cerebral vascular accident/stroke (< 6 months prior to enrollment), myocardial infarction (< 6 months prior to enrollment), unstable angina, congestive heart failure (New York Heart Association Classification Class II), or serious cardiac arrhythmia requiring medication (except for atrial fibrillation that is well controlled with antiarrhythmic medication).	EX10A00 EX10A03	Clinically significant (i.e., active) cardiovascular disease: cerebral vascular accident/stroke (< 6 months prior to enrollment), myocardial infarction (< 6 months prior to enrollment), unstable angina, congestive heart failure ($\geq$ New York Heart Association Classification Class II), or serious cardiac arrhythmia requiring medication.	JB100 criterion is more restrictive as it does not allow controlled atrial fibrillation

	7	Active infection 48 hours before randomization requiring systemic therapy.	EX11A00	Active infection before randomization requiring systemic therapy.	The difference between these two criteria is minimal so they are considered comparable.
	8	Known severe hypersensitivity reactions to monoclonal antibodies (Grade ≥ 3), any history of anaphylaxis, or uncontrolled asthma (i.e., 3 or more features of asthma symptom control per the Global Initiative for Asthma 2015).	EX12A00	Known severe hypersensitivity reactions to monoclonal antibodies (Grade ≥ 3), any history of anaphylaxis, or uncontrolled asthma (i.e., 3 or more features of asthma symptom control per the Global Initiative for Asthma 2015).	
	9	Known prior or suspected hypersensitivity to study drugs or any component in their formulations.	EX13A00	Known prior or suspected hypersensitivity to study drugs or any component in their formulations.	
	10	Diagnosis of prior immunodeficiency or organ transplant requiring immunosuppressive therapy.	EX15A02	Diagnosis of prior immunodeficiency or organ transplant requiring immunosuppressive therapy.	
	11	Pregnant female participants; breastfeeding female participants; and female participants of childbearing potential who are unwilling or unable to use a highly effective method of contraception as outlined in the protocol for the duration of the study and for at least 30 days after the last dose of investigational product.	EX19A00	Pregnant female participants; breastfeeding female participants; and female participants of childbearing potential who are unwilling or unable to use a highly effective method of contraception as outlined in the protocol for the duration of the study and for at least 30 days after the last dose of investigational product.	
	12	Other severe acute or chronic medical conditions including but not limited to inflammatory bowel disease (colitis, Crohn's disease) or gastrointestinal (GI) perforation within 6 months of enrollment, pneumonitis, and pulmonary fibrosis; psychiatric condition including recent (within the past year) or active suicidal ideation or behavior; or laboratory	EX20A00	Other severe acute or chronic medical conditions including but not limited to colitis, inflammatory bowel disease, pneumonitis, and pulmonary fibrosis; psychiatric condition including recent (within the past year) or active suicidal ideation or behavior; or laboratory abnormality that may increase the risk associated with study participation or study treatment	

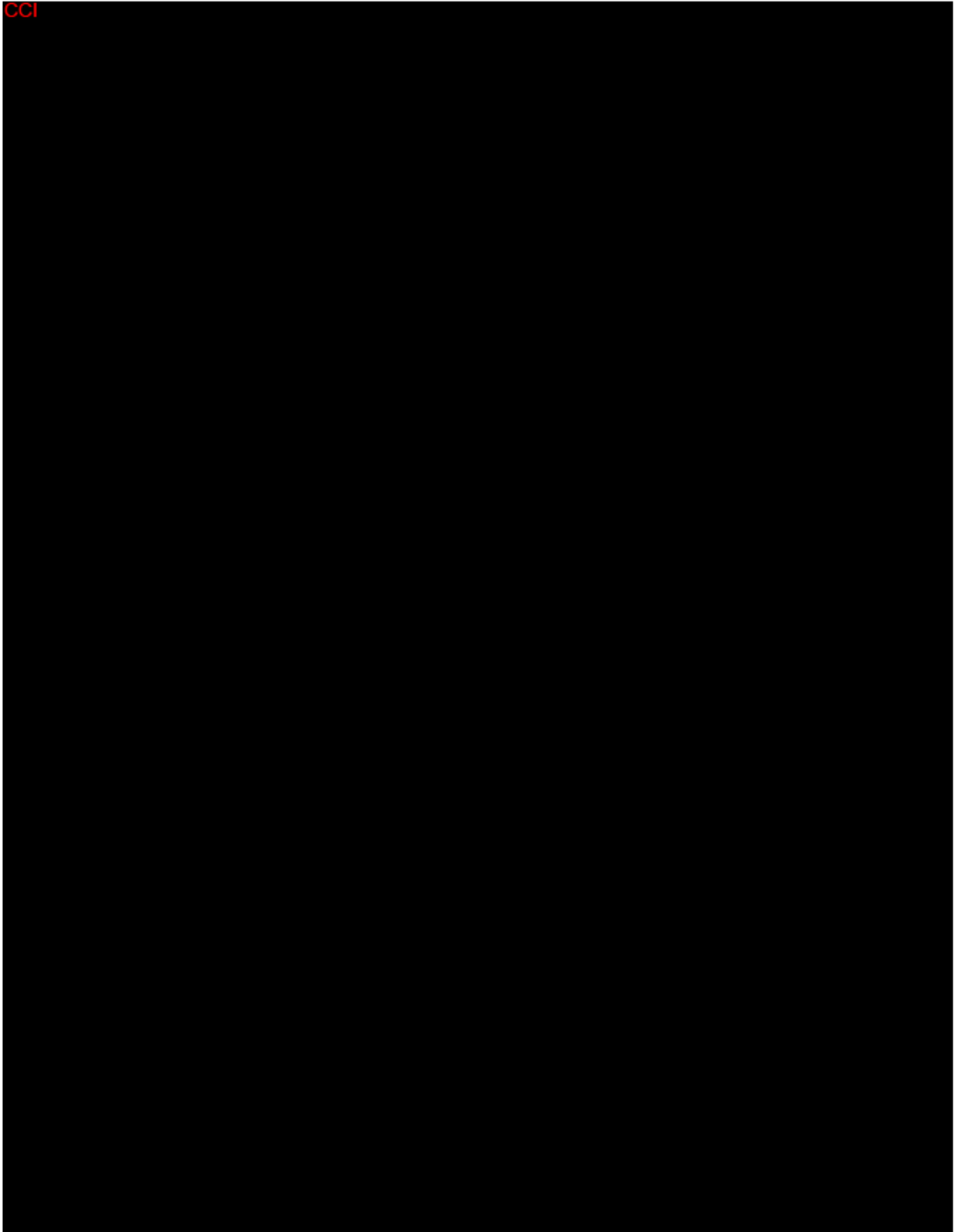
		abnormality that may increase the risk associated with study participation or study treatment administration or may interfere with the interpretation of study results and, in the judgment of the Investigator, would make the participants inappropriate for entry into this study.		administration or may interfere with the interpretation of study results and, in the judgment of the investigator, would make the patient inappropriate for entry into this study.	
Prior/Concomitant Therapy	13	Prior adjuvant or neoadjuvant systemic therapy within 12 months of randomization.	EX02A00 EX02A03	Prior adjuvant or neoadjuvant systemic therapy within 12 months of randomization.	Two exclusion codes were used but the exclusion criterion was not changed.
	14	Prior immunotherapy with IL-2, IL-15, IFN- α , or an anti PD 1, anti PD-L1, anti PD-L2, anti CD137, or CTLA-4 antibody (including ipilimumab), anti TROP2, anti-TIGIT, any other antibody or drug specifically targeting T cell costimulation or immune checkpoint pathways, or any of the investigational drugs used in combination with avelumab.	EX03A00	Prior immunotherapy with IL-2, IFN- α , or an anti-PD-1, anti-PD-L1, anti-PD-L2, anti-CD137, or CTLA-4 antibody (including ipilimumab), or any other antibody or drug specifically targeting T-cell co-stimulation or immune checkpoint pathways.	
	15	Major surgery $\leq$ 4 weeks or major radiation therapy $\leq$ 2 weeks prior to randomization. Prior palliative (pain management or hemostatic) radiotherapy is permitted, provided it has been completed at least 48 hours prior to randomization.	EX04A00 EX04A03	Major surgery $\leq$ 4 weeks or major radiation therapy $\leq$ 2 weeks prior to randomization. Prior palliative (pain management or hemostatic) radiotherapy is permitted, provided it has been completed at least 48 hours prior to randomization.	Criterion became less restrictive while the JBM study was ongoing.
	16	Current or prior use of immunosuppressive medication within 7 days prior to randomization, EXCEPT the following: Intranasal, inhaled, topical steroids, or local steroid injections (e.g., intra-articular injection); Systemic corticosteroids at physiologic doses $\leq$ 10 mg/day of prednisone or equivalent;	EX14A00	Current or prior use of immunosuppressive medication within 7 days prior to randomization, EXCEPT the following: Intranasal, inhaled, topical steroids, or local steroid injections (e.g., intra-articular injection); Systemic corticosteroids at physiologic doses $\leq$ 10 mg/day of prednisone or equivalent;	

		Steroids as premedication for hypersensitivity reactions (e.g., CT scan premedication).		Steroids as premedication for hypersensitivity reactions (e.g., CT scan premedication).	
	17	Vaccination within 4 weeks of the first dose of study treatment and while on trial is prohibited except for administration of inactivated vaccines (for example, inactivated influenza vaccines) and replication-deficient coronavirus vaccines approved or authorized by local Health Authorities.	EX17A00	Vaccination within 4 weeks of the first dose of study treatment and while on trial is prohibited except for administration of inactivate vaccines (for example, inactivated influenza vaccines).	
Prior/Concurrent Clinical Study Experience	18	Participation in other studies involving investigational drug(s) within 4 weeks prior to randomization. Observational studies are permitted.	EX08A00	Participation in other studies involving investigational drug(s) within 4 weeks prior to randomization. Observational studies are permitted.	
Other Exclusions	19	Participants who are investigational site staff members directly involved in the conduct of the study and their family members, site staff members otherwise supervised by the Investigator, or participants who are Merck employees directly involved in the conduct of the study.	EX18A00	Patients who are investigational site staff members directly involved in the conduct of the study and their family members, site staff members otherwise supervised by the investigator, or patients who are Pfizer employees directly involved in the conduct of the study.	

Note: For inclusion criteria, differences are acceptable as long as JB100 is more restrictive than JBM, for exclusion criteria JBM needs to be more restrictive than JB100.

a: The code presented is the one assigned to the variable IETESTCD in the SDTM dataset IE reporting all the inclusion and exclusion criteria that were not met by each screened participant in the JB100 study.

CCI



<u>Subscale</u>	<u>Item Code</u>	<u>Reverse item?</u>	<u>Item response</u>	<u>Item Score</u>
FBISI-DRS-E (Disease Related Symptoms-Emotional)	GE6	4	-	_____ = _____
	GE1	4	-	_____ = _____
Sum individual item scores: _____				
Multiply by 2: _____				
Divide by number of items answered: _____				=FBISI-DRS-E score
Score range: 0-8				

FBISI-TSE (Treatment Side Effects)	GP2	4	-	_____ = _____
	GP1	4	-	_____ = _____
	GP6	4	-	_____ = _____
	C3	0	+	_____ = _____
	GP5	4	-	_____ = _____

Each item should be treated individually with an emphasis on GP5

FBISI-F/WB (Function/ Well-Being)	GF3	0	+	_____ = _____
	GF7	0	+	_____ = _____
Sum individual item scores: _____				
Multiply by 2: _____				
Divide by number of items answered: _____				=FBISI-F/WB score
Score range: 0-8				

*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.

FBISI-18 scoring template, 2008

Appendix 4 – NCI-CTC Gradable and Non-Gradable Safety Laboratory Test Parameters and Direction(s) of Abnormality

NCI-CTC gradable parameters

Laboratory Assessment	Parameters	Name in NCI-CTC	Direction(s) of abnormality
Hematology	Hemoglobin ^a	Anemia / Hemoglobin increased	Low / High
	Eosinophils	Eosinophilia	High
	Leukocytes (WBC) ^a	White blood cell decreased / Leukocytosis	Low / High
	Neutrophils	Neutrophil count decreased	Low
	Lymphocytes	Lymphocyte count decreased / increased	Low / High
	Platelets ^a	Platelet count decreased	Low
Biochemistry	Alanine Aminotransferase (ALT) ^a	Alanine Aminotransferase increased	High
	Aspartate Aminotransferase (AST) ^a	Aspartate Aminotransferase increased	High
	Alkaline Phosphatase (ALP)	Alkaline Phosphatase increased	High
	Total Bilirubin ^a	Blood bilirubin increased	High
	Amylase	Serum amylase increased	High
	Lipase	Lipase increased	High
	Creatinine ^a	Creatinine increased	High
	Sodium	Hyponatremia / Hypernatremia	Low / High
	Potassium	Hypokalemia / Hyperkalemia	Low / High
	Calcium Corrected	Hypocalcemia / Hypercalcemia	Low / High
	Glucose ^a	Hypoglycemia	Low
	Magnesium	Hypomagnesemia / Hypermagnesemia	Low / High
	Creatine Kinase (also known as creatine phosphokinase) ^a	CPK increased	High
Coagulation	Prothrombin Intl. Normalized Ratio (INR) ^a	INR increased	High

Note: parameters with both Low and High directions of abnormality are going to be split. For example, Potassium is going to be split in Potassium Low and Potassium High.

^a Parameters to be presented in boxplots.

NCI-CTC non-gradable parameters

Laboratory Assessment	Parameters	Direction(s) of abnormality
Hematology	Hematocrit	Low / High
	Mean corpuscular volume (MCV)	Low
	Mean corpuscular hemoglobin (MCH)	Low
	Erythrocytes (RBC) ^a	Low / High
	Reticulocytes	Low / High
	Reticulocytes/Erythrocytes	Low / High
	Neutrophils/Leukocytes (%)	Low / High
	Lymphocytes/Leukocytes (%)	Low / High
	Basophils	High
	Basophils/Leukocytes (%)	High
	Eosinophils/Leukocytes (%)	High
	Monocytes	High
	Monocytes/Leukocytes (%)	High
Biochemistry	Phosphate (Phosphorous)	Low
	Chloride	Low / High
	Urea	High
	Urea Nitrogen (BUN)	High
Hormonal	Thyroid-stimulating hormone (TSH) ^a	Low / High
	Thyroxine (T4) ^a	Low / High
	Free Thyroxine (Free T4) ^a	Low / High

Note: parameters with both Low and High directions of abnormality are going to be split. For example, Chloride is going to be split in Chloride Low and Chloride High.

^a Parameters to be presented in the boxplots.