Nektar Therapeutics

STATISTICAL ANALYSIS PLAN

A PHASE 3 RANDOMIZED OPEN-LABEL STUDY TO COMPARE NKTR-214 COMBINED WITH NIVOLUMAB TO THE INVESTIGATOR'S CHOICE OF SUNITINIB OR CABOZANTINIB IN PATIENTS WITH PREVIOUSLY UNTREATED ADVANCED RENAL CELL CARCINOMA

Protocol Number:	17-214-09, Amendment 3.0
SAP Version:	Final 2.0
Date:	January 24, 2022

[redacted]

PREPARED BY:

[redacted]

[redacted]

TABLE OF CONTENTS

[redacted]

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	adverse event
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AST	aspartate aminotransferase
ATC	anthracycline, taxane, capecitabine
ATCL	Anatomical, Therapeutic Chemical Level classification
BCBM	breast cancer brain metastases
BFI	Brief Fatigue Inventory
BICR	blinded independent central review
BMI	body mass index
BN-20	quality of life assessment specific to brain neoplasms
BSA	body surface area
CBR	clinical benefit rate
CMH	Cochran-Mantel-Haenszel
CNS	central nervous system
CR	complete response
DoR	duration of response
DMC	data monitoring committee
eCRF	electronic case report form
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organization for Research and Treatment of Cancer
EQ-5D-3L™	EuroQol 5D 3 Level Version
ESMO-MCBS	European Society for Medical Oncology magnitude of clinical benefit scale
GCP	Good Clinical Practice
GPA	Graded Prognostic Assessment
HER	human epidermal growth factor receptor
HR	hazard ratio
HRQoL	health-related quality of life
IPFI	initial progression-free interval
IRT	interactive response technology
ITT	intent-to-treat
IV	intravenous
KM	Kaplan-Meier
KPS	Karnofsky Performance Status
LBA	Ligand binding assay

Abbreviation	Definition
LDH	lactate dehydrogenase
LR	locally recurrent
MBC	metastatic breast cancer
MedDRA	Medical Dictionary for Regulatory Activities
MID	minimal important difference
MMRM	Repeated Measures Linear Mixed Effects Model
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NE	not evaluable
ORR	objective response rate
OS	overall survival
PD	progressive disease
PE	physical examination
PEG	polyethylene glycol
PFS	progression-free survival
PK	pharmacokinetic(s)
PR	partial response
Q21d	once every 21 days
QLQ-C30	Quality of Life Core 30
QoL	quality of life
RANO-BM	Response Assessment in Neuro-Oncology—Brain Metastases
RECIST	Response Evaluation Criteria in Solid Tumors
RS	raw score
SAE	serious adverse event
SD	stable disease
SN38	7-ethyl-10-hydroxy-camptothecin; the active metabolite of irinotecan
SOC	system organ class
SRS	stereotactic radiosurgery
TEAE	Treatment-emergent adverse event
TNBC	triple-negative breast cancer
TPC	Treatment of Physician’s Choice
UGT1A1	uridine diphosphate-glucoronosyl transferase 1A1
WHO-DDE	World Health Organization Drug Dictionary Enhanced

1.0 ADMINISTRATIVE STRUCTURE

This study will be managed via partnership between Nektar Therapeutics and a contract research organization (CRO). Central clinical laboratories will be used for processing of safety specimens, biomarkers, immunogenicity, pharmacokinetic (PK), and pharmacodynamics (PD) samples. An interactive response technology (IRT) service provider will manage the randomization system, study drug, and comparator distribution and inventory management. Data for this trial will be entered into an Electronic Data Capture (EDC) system, using a Medidata Rave platform. An independent data monitoring committee (DMC) is established to review the interim efficacy and safety data on a periodic basis. Response and progression will be determined by blinded independent central review (BICR) using modified Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1; see modified RECIST 1.1 criteria in Appendix 8 of the clinical study protocol).

2.0 INTRODUCTION

This document describes the planned statistical analyses of the efficacy, safety, PK, PD, biomarker and immunogenicity data captured according to Nektar Therapeutics Protocol 17-214-09 "A Phase 3 Randomized Open-Label Study to Compare NKTR-214 Combined with Nivolumab to the Investigator's Choice of Sunitinib or Cabozantinib in Patients with Previously Untreated Advanced Renal Cell Carcinoma" version amendment 3.0 dated 19 October 2021.

This Phase 3 study is conducted in accordance with the protocol, Good Clinical Practice (GCP), and all applicable regulatory requirements.

3.0 STUDY OBJECTIVES

3.1 Primary Objectives

- To compare the objective response rates (ORR) by blinded independent central review (BICR) assessment using modified Response Evaluation Criteria in Solid Tumors (mRECIST) 1.1 of NKTR-214 combined with nivolumab to that of tyrosine kinase inhibitor (TKI) monotherapy (sunitinib or cabozantinib) in International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) intermediate- or poor-risk patients with previously untreated advanced renal cell carcinoma (RCC)
- To compare the ORR by BICR using mRECIST 1.1 of NKTR-214 combined with nivolumab to that of TKI monotherapy (sunitinib or cabozantinib) in IMDC all-risk patients with previously untreated advanced RCC
- To compare the overall survival (OS) of NKTR-214 combined with nivolumab to that of TKI monotherapy (sunitinib or cabozantinib) in IMDC intermediate- or poor- risk patients with previously untreated advanced RCC
- To compare the OS of NKTR-214 combined with nivolumab to that of TKI monotherapy (sunitinib or cabozantinib) in IMDC all-risk patients with previously untreated advanced RCC

3.2 Secondary Objectives

The key secondary objectives are:

- To compare the progression-free survival (PFS) by BICR using mRECIST 1.1 of NKTR-214 combined with nivolumab to TKI monotherapy (sunitinib or cabozantinib) in IMDC intermediate- or poor- patients with previously untreated advanced RCC
- To compare the PFS by BICR using mRECIST 1.1 of NKTR-214 combined with nivolumab to TKI monotherapy (sunitinib or cabozantinib) in IMDC all-risk patients with previously untreated advanced RCC

The other secondary objectives are:

- To estimate the incidence of adverse events (AEs) of NKTR-214 combined with nivolumab versus TKI monotherapy (sunitinib or cabozantinib) in patients with previously untreated advanced RCC
- To evaluate whether programmed cell death ligand 1 (PD-L1) expression ($< 1\%$ vs $\geq 1\%$) using the PD-L1 immunohistochemistry (IHC) 28-8 pharmDx assay is a predictive biomarker for ORR, PFS, and OS in patients with previously untreated advanced RCC
- To characterize changes in cancer-related symptoms and quality-of-life in patients with previously untreated advanced RCC using the National Comprehensive Cancer Network/Functional Assessment of Cancer Therapy (NCCN/FACT) Symptom Index for Kidney Cancer (FKSI-19)

4.0 STUDY ENDPOINTS

4.1 Primary Efficacy Endpoints

- ORR, defined as the percentage of patients with a confirmed best overall response of complete response (CR) or partial response (PR) per mRECIST 1.1 by BICR.
- OS, defined as time from randomization to the date of death from any cause. Patients who do not have a date of death will be censored at their last known alive date.

4.2 Key Secondary Efficacy Endpoint

- PFS, defined as time from randomization to the first date of documented tumor progression per mRECIST 1.1 by BICR or death due to any cause, whichever comes first.

4.3 Other Efficacy Endpoints

- PFS2 (PFS after the next line of treatment), defined as time from randomization to objectively documented progression, per investigator assessment, after the next line of therapy or to death from any cause, whichever occurs first.
- Duration of response (DOR), defined as the time between the date of first confirmed documented response (CR or PR) to the date of the first documented tumor progression as determined per mRECIST 1.1 by the BICR, or death due to any cause, whichever occurs first.
- Time to response (TTR), defined as the time from randomization to the date of the first confirmed documented response (CR or PR) per mRECIST 1.1 by BICR.
- Clinical benefit rate (CBR), defined as defined as the number of patients with CR, PR, or SD per mRECIST 1.1 by BICR

5.0 OVERALL STUDY DESIGN AND PLAN

5.1 Study Design

This is a multicenter, randomized, open label, Phase 3 study that will evaluate the efficacy and safety of NKTR-214 combined with nivolumab compared with the Investigator's choice of a TKI, either sunitinib or cabozantinib, in patients with previously untreated advanced RCC.

The treatment period of the study is divided into multiple treatment cycles with associated evaluations and procedures. One cycle of treatment is defined as 6 weeks. Results of the assessments must be reviewed and documented before administering the scheduled dose of study drugs. In certain circumstances, patients with progressive disease (PD) per mRECIST 1.1 but with otherwise stable or improved performance and clinical status may continue to be treated in the event of a perceived benefit per Investigator.

This study is divided into a Screening period, a Treatment period, and a Long-Term Follow-Up period, which will continue until withdrawal of consent, death, lost to follow-up, or study termination by the Sponsor.

Approximately 600 patients (including at least 500 patients with IMDC intermediate- or poor-risk and approximately 100 patients with IMDC favorable risk) with previously untreated advanced RCC will be randomized in a 1:1 ratio to the two treatment arms below:

- Arm A: NKTR-214 0.006 mg/kg IV Q3W combined with nivolumab 360 mg IV Q3W
- Arm B: Investigator's choice of either one of the following treatments
  - Sunitinib 50 mg po once daily for 4 weeks followed by 2 weeks off

    OR
  - Cabozantinib 60 mg po once daily

Randomization will be stratified by:

- IMDC prognostic score (0 [favorable risk] versus 1-2 [intermediate risk] versus 3-6 [poor risk]); laboratory data used to calculate IMDC score are based on central laboratory results)
- TKI choice (sunitinib versus cabozantinib)

5.2 Study Medications

Patients should begin study treatment within 5 calendar days of randomization.

Treatment will continue until progression, unacceptable toxicity, withdrawal of consent, completion of 2 years (Arm A only) of treatment, the study ends, or other protocol specified reasons, whichever occurs first.

5.2.1 Arm A: NKTR-214 Combined with Nivolumab

Each patient's NKTR-214 dose will be determined by the patient's weight in kilograms, which will be determined on the day of each NKTR-214 dosing. If the patient's weight is within 10% of the Cycle 1 Day 1 weight, the study drug dose does not need to be recalculated depending on institutional guidelines/preference. Patients should receive NKTR-214 at a starting dose of 0.006 mg/kg on Day 1 and Day 22 of each treatment cycle.

NKTR-214 will be administered before nivolumab via IV infusion over 30 (± 5) minutes Q3W. Nivolumab administration should start at least 30 minutes after completion of the NKTR-214 infusion. Patients should receive nivolumab at a flat-dose of 360 mg as a 30-minute infusion on Day 1 and Day 22 of each treatment cycle.

For patients randomized to Arm A, in the event that one of the study drugs is stopped due to toxicities, the other drug can be continued as a monotherapy if the toxicities are considered related to only one of the study drugs, once the criteria to resume are met.

5.2.2 Arm B: Treatment of Investigator's Choice

Patients randomized to TKI will receive a widely used standard-of-care agent: sunitinib or cabozantinib. Sunitinib is a vascular endothelial growth factor (VEGF) receptor TKI that is approved and recommended for the treatment of advanced RCC across prognostic groups. The starting dose of sunitinib is 50 mg, and it is administered po once daily for 4 weeks followed by 2 weeks off (6-week cycle). Cabozantinib is a small molecule inhibitor of the tyrosine kinases c-Met, AXL, and VEGFR2, and has been shown to reduce tumor growth, metastasis, and angiogenesis. Cabozantinib has been evaluated in both first-line and second-line settings in advanced RCC. The starting dose of cabozantinib is 60 mg (40 mg for patients with moderate hepatic impairment [Child-Pugh score B]) and it is administered po once daily.

5.3 Pharmacokinetic, Pharmacodynamic, [redacted] Study Measurements and Endpoints

PK assessments are only performed for Arm A. Blood samples for the assessment of NKTR-214 and nivolumab concentrations will be collected and analyzed.

[redacted]

5.3.1 Blood Sampling Times

Blood samples for PK, immunogenicity, and [redacted]/pharmacodynamic assessments of NKTR-214 and nivolumab will be collected as described in Protocol Amendment 2.0 Table 6.

5.3.2 Tumor Biopsy Times

Tumor biopsies or archival tissues will be collected as described in Table 1. The samples for the formalin-fixed paraffin-embedded block or unstained slides sectioned from a block will be collected within 3 months prior to randomization (minimum of 8 slides, preferably 20 slides).

Table 1: Tumor Tissue Collection Schedule

Study Day of Sample Collection (1 cycle= every 6 weeks)[a]	**Archival or Fresh Tumor Tissue in FFPE**
Screening	X
Cycle 1 Days 14-21	X[b]
Disease Progression	X[b]

Source: Table 6 of Study protocol

[redacted]

b. Optional but highly recommended

5.3.3 Pharmacokinetic Analysis Set (PKAS)

The pharmacokinetic analysis set (PKAS) will consist of 4 different analyses. The PKAS includes all patients who received at least 1 dose of NKTR-214 and at least 1 measurable concentration of:

- NKTR-214-RC (related cytokine) PKAS: NKTR-214-RC
- NKTR-214-AC (active cytokine) PKAS: NKTR-214-AC
- Total-PEG PKAS: Total-PEG
- Nivolumab PKAS: Nivolumab

5.3.4 Pharmacokinetic Measurements and Endpoints

NKTR-214-RC and Total-PEG will be analyzed using fully validated ligand binding assays (LBA). NKTR-214-AC will be analyzed using a qualified LBA.

Nivolumab will be assayed using a fully validated LBA method. These assays are summarized in Table 2.

Table 2: NKTR-214 and Nivolumab Pharmacokinetic Measurements

Assay	Matrix	Assay Format	Assay Type	Analytes Captured
NKTR-214-RC	Plasma	LBA	Validated	Mixture of compounds containing IL-2 independent of PEG conjugation status
NKTR-214-AC	Plasma	LBA	Qualified	Mixture of 2-PEG-IL-2, 1-PEG-IL-2 and free IL-2
Total-PEG	Plasma	LBA	Validated	PEG, independent of conjugation status to IL-2
Nivolumab	Serum	LBA	Validated	Nivolumab

Abbreviations: LBA = ligand binding assay

The analysis methods described in Section 12.1 will be used to estimate the Cycle 1 PK parameters for NKTR-214-RC, NKTR-214-AC, and Total-PEG for individual patients based on observed concentration values. Cycle 3 PK parameters will also be estimated if data allows.

[redacted]

5.4 Immunogenicity Measurements and Endpoints

Immunogenicity assessments are only performed for Arm A. Blood samples for the assessment of the anti-drug antibodies of NKTR-214 and nivolumab will be collected and measured as described in Table 6 of the protocol.

Validated methods to detect anti-NKTR-214/anti-PEG, anti-IL-2 and anti-nivolumab anti-drug antibodies (ADA) will be used to analyze immunogenicity samples. Immunogenicity sample testing will be done in tiers as per the 2019 FDA guidance (FDA 2019). Samples will be first tested with screening electrochemiluminescence assays (ECLAs). Putative positive samples for anti-NKTR-214, anti-IL-2 and anti-nivolumab ADA will then be analyzed in competition ECLAs to confirm positivity. Confirmed anti-NKTR-214 ADA-positive samples will be tested further in a polyethylene glycol (PEG) immuno-competition assay to determine the antibody specificity of the reactivity to the PEG or non-PEG (IL-2, linker) moiety of NKTR-214. Confirmed positive samples from each assay (anti-nivolumab, anti-NKTR-214 and anti-IL-2) will then be tested to obtain a titer Samples confirmed to be positive for anti-NKTR-214/anti-PEG, anti-IL-2 antibodies will be tested for neutralizing activity for IL-2 using a validated cell-based assay. Anti-nivolumab antibodies positive samples will also be tested for neutralizing activity for nivolumab using a validated cell-based assay.

6.0 STATISTICAL CONSIDERATIONS

6.1 General Considerations

Although this is an open label study, analyses for treatment comparisons prior to final database lock will not be performed except for the pre-planned interim safety and efficacy analyses as described in Section 11.0 of this document and in the DMC charter.

Blinding for aggregated treatment information during the trial and data unblinding for interim analysis and after final database lock are provided in the Aggregate Data Handling Plan.

Summary statistics for continuous variables will include the mean, standard deviation, median, minimum, and maximum. The mean and median will be presented to one decimal place beyond which the data were captured. The standard deviation will be presented to two decimal places beyond which the data were captured. The minimum and maximum will be presented to the precision with which the data were captured.

Categorical variables will be presented as frequency counts and percentages. A row or column denoted 'Missing' will be included in count tabulations where necessary to account for dropouts and missing values. Percentages will be rounded to 1 decimal place and the percent will be suppressed when the count is zero. The denominator will be the number of patients in the population of interest unless otherwise noted.

Time-to-event variables will be analyzed using the Kaplan-Meier (KM) method. The KM estimates for 25^{th}, 50^{th} (median) and 75^{th} percentiles with associated 2-sided 95% confidence interval using Greenwood's formula (Greenwood, 1926), as well as the percentage of censored observations will be presented. All time-to-event variables will be plotted using the KM method.

Data listings will be created to support each table and to present all data. Data listings will be presented by treatment group and patient number.

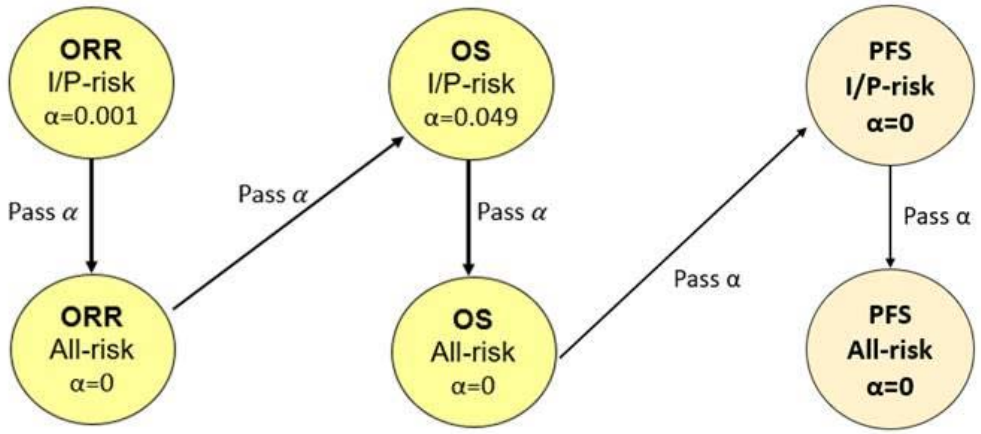
6.2 Determination of Sample Size

The sample size of the study accounts for the following primary efficacy endpoints: ORR in IMDC intermediate- or poor-risk and IMDC all-risk populations, and OS in IMDC intermediate- or poor-risk and IMDC all-risk populations. The overall alpha for this study is 0.05, which is split with 0.001 to evaluate ORR and 0.049 to evaluate OS. All alpha levels are two-sided. Statistical testing for the IMDC all-risk population will be gated by the corresponding tests in the IMDC intermediate- or poor- risk population for both ORR and OS. That is, ORR in the IMDC all-risk population will not be tested unless ORR in IMDC intermediate- or poor-risk population is statistically significant; and similarly, OS in the IMDC all-risk population will not be tested unless OS in IMDC intermediate- or poor- risk population is statistically significant.

In addition, a fallback approach will be used to reallocate alpha from ORR to OS. If ORR is significant in the IMDC intermediate- or poor-risk population and the IMDC all-risk population,

the alpha of 0.001 will be passed to OS. OS in the IMDC intermediate- or poor-risk population can then be tested at the 0.05 (instead of 0.049) level. If it is significant, the alpha of 0.05 will be passed to OS in the IMDC all-risk population. The testing procedure of ORR and OS is described in Figure 1. The total sample size is 600 patients. It is estimated that at least 500 IMDC intermediate- or poor-risk and approximately 100 IMDC favorable-risk, previously untreated advanced RCC patients will be randomized into the two treatment arms (NKTR-214 combined with nivolumab vs. TKIs) in a 1:1 ratio.

Figure 1: Testing Procedure for Primary Endpoints and Key Secondary Endpoints



Abbreviations: All-risk = IMDC all-risk population; I/P risk = IMDC intermediate- or poor-risk population; ORR = objective response rate; OS = overall survival; PFS = progression-free survival.

6.2.1 Sample Size Justification for ORR

Given a two-sided alpha of 0.001, 500 patients in the IMDC intermediate- or poor-risk population will provide approximately 82% power to detect an 18% difference in ORR between the 2 treatment arms using a Chi-square test, assuming an ORR of 27% for TKIs. If the observed ORR difference is greater than 13.9%, the result will be statistically significant. Six hundred patients in the IMDC all-risk population will provide approximately 89% power to detect an 18% difference in ORR between 2 treatment arms using a Chi-square test, assuming an ORR of 32% for TKIs. If the observed ORR difference is greater than 13.2%, the result will be statistically significant. If ORR is significant in IMDC intermediate- or poor-risk population, the alpha of 0.001 will be passed to ORR in IMDC all-risk population.

This ORR analysis will occur at the time of the interim analysis of OS. Based on the average accrual rate of 28 patients per month, at the time of OS interim analysis, it is estimated that 600 patients in the IMDC all-risk population would have been enrolled and followed for 6 months of which at least 500 patients will be IMDC intermediate- or poor-risk.

6.2.2 Sample Size Justification for OS

The study is powered to detect an improvement in OS for NKTR-214 combined with nivolumab as compared to TKIs. Due to the long duration of OS and potential subsequent non-protocol therapies, there may be confounding issues for the OS endpoint. Thus, it is critically important to use interim OS results to efficiently power the OS endpoint and conduct the final analysis as early as possible using an adaptive approach. The interim analysis will be triggered when at least 156 OS events in IMDC intermediate- or poor- risk population are observed. All primary endpoints will be tested at this interim analysis. Two parallel sequential tests will be conducted: (1) ORR in IMDC intermediate- or poor- risk and IMDC all-risk population at $\alpha = 0.001$ and (2) OS in IMDC intermediate- or poor- risk and IMDC all-risk population at $\alpha = 0.01$. Within (1), ORR in IMDC all-risk population will only be tested at 0.001 if ORR in IMDC intermediate or poor risk is statistically significant at 0.001. Within (2), OS in IMDC all-risk population will only be tested at 0.01 if OS in IMDC intermediate or poor risk is statistically significant at 0.01.

If OS in IMDC intermediate- or poor-risk is not statistically significant, conditional power will be calculated based on this population. The adaptive method in Liu and Hu (2016), and Mehta and Pocock (2011) will be used to determine the number of OS events in IMDC intermediate- or poor-risk population to target approximately 85% to 90% conditional power for the final analysis. The final analysis will be triggered when this re-estimated number of OS events in IMDC intermediate- or poor- risk is reached. The test of OS in IMDC all-risk population will also be performed at the time of the final analysis regardless of number of events observed in this group.

If OS in IMDC intermediate- or poor-risk is statistically significant, but not in the IMDC all-risk population, conditional power will be calculated for the IMDC all-risk population. The adaptive method in Liu and Hu (2016), and Mehta and Pocock (2011) will be used to determine the number of OS events in IMDC all-risk population to target approximately 85% to 90% conditional power for the final analysis.

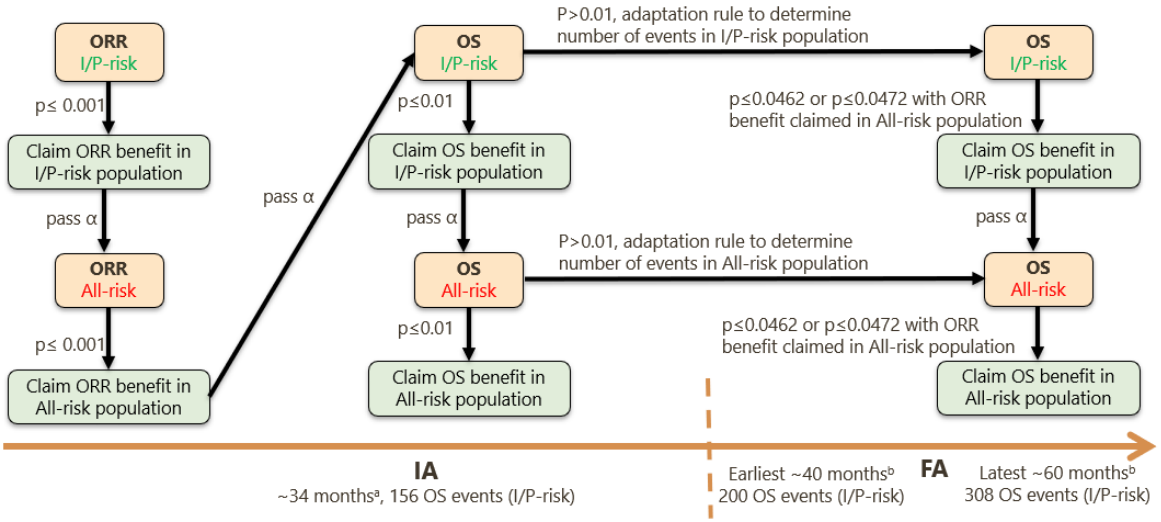
The minimum and maximum number of OS events in IMDC intermediate- or poor-risk population for the final analysis are 200 and 308, respectively. The minimum event size is calculated to achieve 90% power with alpha=0.049 (two-sided) assuming an exponential distribution in each treatment arm and a HR = 0.63 with median OS for the TKI arm of 26 months in IMDC intermediate- or poor- risk population. The maximum event size is determined based on the minimum clinically meaningful effect. When the minimum event size of 200 is reached in the IMDC intermediate- or poor-risk population, approximately 211 events are expected to be observed in the IMDC all-risk population, which gives approximately 80% power with alpha=0.049 (two-sided) assuming an exponential distribution in each treatment arm and a HR = 0.68 with median OS for the TKI arm of 32.9 months in the IMDC all-risk population.

The detailed sample size adaptation rules based on conditional power will be described in an appendix of the DMC Charter, which will not be accessible to the study team or sites to prevent introduction of any potential bias; however, the operational procedures will be described in the

DMC charter, which is accessible to the study team. The key secondary endpoint of PFS in the IMDC intermediate- or poor-risk population and the IMDC all-risk population will be tested hierarchically given the statistical significance of OS.

Figure 2 provides the study flow and decision-making time points.

Figure 2: Study Flow and Decision-Making Time Points



IA=interim analysis, FA=final analysis, I/P-risk=IMDC intermediate or poor risk, All-risk=IMDC all risk
a. The timeline for IA is based on accrual and the OS assumptions and will be triggered based on the number of OS events in I/P-risk population.
b. The timeline for FA will be triggered based on the number of OS events in I/P-risk or all-risk population.

6.3 Analysis Populations

Intent-to-Treat (ITT) IMDC Intermediate- or Poor-risk Population: All patients who are randomized and are classified as intermediate- or poor risk according to the IMDC prognostic score entered in the IRT system at randomization. Patients are grouped by the treatment to which they are randomized. It is abbreviated as ITT I/P-risk Population in the document.

ITT IMDC All-risk Population: All patients who are randomized. Patients are grouped by the treatment to which they were randomized. It is abbreviated as ITT All-risk Population in the document.

Safety IMDC Intermediate- or Poor-risk Population: All patients who receive at least 1 dose (or partial dose) of any study drug and are classified as the intermediate- or poor-risk according to the IMDC prognostic score entered in the IRT system at randomization. Patients are grouped according to the treatment they actually received. It is abbreviated as Safety I/P-risk Population in the document.

Safety IMDC All-risk Population: All patients who receive at least 1 dose (or partial dose) of study drug. Patients are grouped according to the treatment they actually received. It is abbreviated as Safety All-risk Population in the document.

Objective Response Rate (ORR) IMDC Intermediate- or Poor-risk Population: All randomized patients classified as the intermediate- or poor-risk according to the IMDC prognostic score who have at least 6 months of follow-up at the time point of the ORR analysis. A patient's follow-up time is defined here as the time between randomization date and clinical data cutoff date regardless of patient disposition (i.e., ITT principle). Patients are grouped by the treatment to which they were randomized. This analysis population will be equivalent to the ITT IMDC Intermediate- or Poor-risk Population if the clinical data cutoff date is at least 6 months after the date that the last patient was randomized. It is abbreviated as ORR I/P-risk Population in the document.

ORR IMDC All-risk Population: All randomized patients who have at least 6 months of follow-up at the time point of the ORR analysis. A patient's follow-up time is defined here as the time between randomization date and clinical data cutoff date regardless of patient disposition (i.e., intent-to-treat principle). Patients are grouped by the treatment to which they were randomized. This analysis population will be equivalent to the ITT IMDC All-risk Population if the clinical data cutoff date is at least 6 months after the date that the last patient was randomized. It is abbreviated as ORR All-risk Population in the document.

Pharmacokinetics (PK) Population: The PK Population is defined as all patients in the Safety IMDC All-risk Population who have at least one quantifiable PK concentration for either NKTR-214-RC, NKTR-214-AC, Total-PEG, or nivolumab.

[redacted]

Immunogenicity Population: The Immunogenicity Population is defined as all patients in the Safety IMDC All-risk Population Arm A (i.e., treated with NKTR-214 and nivolumab) who have at least baseline and 1 post-baseline ADA assessment.

6.4 Handling of Missing Data

- For the subsequent anti-cancer therapy start date:
  - If only the day is missing then: if the end date of the subsequent anti-cancer therapy is complete, impute to the minimum of (the last day of the month, end date of subsequent anti-cancer therapy); otherwise impute to the last day of the month.
  - If the month or year is missing, then leave missing.
- For the time from initial diagnosis to randomization and the time from diagnosis of metastatic disease to randomization, partial dates for initial diagnosis and diagnosis of metastatic disease will be imputed as follows:
  - No imputation will be done if the year is missing

  - If the year is before the randomization date, then missing days will be imputed as the first day of the month and missing months will be imputed as July
  - If the year is the current year of the randomization date, then missing days will be imputed as the first day of the month and missing months will be imputed as January
- For the duration of AEs, partial dates for the **start** of the AE will be imputed as follows:
  - Missing day, month, and year should be queried. In case of non-resolution of missing year, then the missing day will be imputed as Cycle 1 Day 1 (C1D1).
  - Start day of AE is missing and the year is same as Cycle 1 Day 1 (C1D1)
    - If the reported month of occurrence of the AE is after the month of the C1D1 dose, then missing day will be imputed as the first day of the month of occurrence of AE.
    - If the reported month of occurrence of AE is the month of C1D1 dose then the missing day will be imputed as the same day as C1D1.
    - If the reported month of the AE start date is before the month of C1D1, then the missing day will be imputed as day 15 of the month of the AE start date.
    - If the month of the AE start date is missing, the missing day will be imputed as the date of C1D1.
  - Start day of the AE is missing and the year is after the year of C1D1
    - Missing day will be imputed as the first day of the month of the occurrence of the AE
    - If the month of the AE start date is missing, the missing day will be imputed as 01 January of the year of the AE start date.
- For the duration of AEs, partially missing dates for the **stop** of the AE will be imputed as follows:
  - If the stop date is completely missing, it should be queried. In case of non-resolution of missing year, no imputation will be performed.
  - If only the day is missing, the earlier of the following will be used as the stop date:
    - the last day of the month for stop of AE
    - the later of (last known alive date, AE start date if imputed)
  - If themonth and day are missing, the earlier of the following will be used as the stop date:
    - December 31st of the year for stop of AE
    - the later of (last known alive date, AE start date if imputed)
- For the determination of prior medication, any medication with a start date prior to C1D1 will be classified as prior medication regardless of the the stop date. Missing or partial dates will be handled as follows:

  - If missing day and/or month of the start date, the medication will be classified as prior unless the month and/or year of the start date is after C1D1

- For determination of concommitant medication, the following will be classified as concomitant medication:
  - Any medication with a start date prior to or on C1D1 and continued after C1D1 (this will be considered as both prior and concomitant medication)
  - Any medication with a start date after C1D1, but prior to the last dose date + 100 days or (date of initiation of subsequent anti-cancer therapy – 1 day), whichever is earlier
- Missing or partial dates for concomitant medication will be handled as follows:
  - If missing the day and/or month of the start date, the medication will not be considered as concomitant if the month and/or year of the start date is after the last dose date + 100 days or (date of initiation of subsequent anti-cancer therapy – 1 day), whichever is earlier
  - If missing the day and/or month of the stop date, the medication will not be considered as concomitant if the month and/or year of the stop date is prior to C1D1
  - A medication with completely missing start and stop dates will be classified as concomitant
- In order to calculate duration of immune modulating medication for management of drug-related select AEs or IMAEs (as defined in Section 8.2.2), missing or partial dates for these medications will be handled as follows:

  Missing or partial start date:

  - If the start date is completely missing, it will not be imputed. Only the number of days that the medication was taken on or after the AE start date will be counted toward the duration of IMM for management of certain AE.
  - If only the year is provided, impute the start date to January 1st of the year,
  - If only the day is missing, impute the start date to the first day of the month.

  Missing or partial stop date:

  - If the stop date is completely missing, impute stop date to the later of (last known alive date, AE start date if imputed)
  - If only the day is missing, the earlier of the following will be used as the stop date:
    - the last day of the month for stop of AE
    - the later of (last known alive date, AE start date if imputed)
  - If month and day are missing, the earlier of the following will be used as the stop date:

  - December 31st of the year for stop of AE
  - the later of (last known alive date, AE start date if imputed)

- For death date, the following conventions will be used:
  - If the death date is completely missing but reason for death is present, it will be imputed as the last known alive date
  - If the the death date is partially missing (missing day only, or missing both day and month), the death date and corresponding last known alive date will be imputed in 2 steps:
    1. Imputed as the 1st of the month (missing day only) or January 1st of the year (missing both day and month)
    2. The imputed death date from the 1st step will be compared to the latest alive date captured from multiple sources (unimputed last known alive date), and the maximum of the two will be considered as the imputed death date and last known alive date

Some laboratory analytes may be reported by range (e.g., gamma-glutamyl transferase [GGT] < 17 U/L) and will be imputed by the corresponding boundary (e.g., 17 U/L).

Handling of missing data for efficacy analyses and HRQoL analyses are described in Section 7.5 and Section 7.6, respectively.

No imputation of other missing data is planned.

6.5 Stratification and Pooling

For stratified analyses, the randomization stratification factors (IMDC prognostic score and TKI choice) entered into the IRT system at the time of randomization will be used, as follows:

- IMDC prognostic score as entered in IRT (0 [favorable risk] vs 1-2 [intermediate risk] vs 3-6 [poor risk])
- TKI choice as entered in IRT (sunitinib vs cabozantinib)

The IMDC prognostic score will also be derived based on data entered in EDC. When deriving IMDC prognostic score, the latest assessment prior to or on the date of randomization will be used (if not available, then the latest assessment before the date and time of first dose of study drug will be used); if the site used local laboratory data for randomization, the local laboratory data collected in EDC cancer history form will be used. Discrepancies between IMDC prognostic score derived vs IMDC prognostic score as entered in IRT system will be summarized.

If the discrepany between the IMDC prognostic score (0 vs 1-2 vs 3-6) derived versus as entered in IRT is ≥ 10%, a sensitivity stratified analyses using strata based upon derived IMDC

prognostic score per EDC and TKI choice will be performed for the primary and key secondary endpoints.

To avoid unstable estimates, if any stratum (defined by the stratification factors for analysis) have fewer than 6 patients (1% of the estimated total number of patients) or there are no events, that stratum will be pooled with the smallest adjacent stratum for analyses. The smallest adjacent stratum is defined as the stratum with the fewest number of patients or events if the former is a tie, and the adjacent stratum is defined as the stratum with one of the two factors at the same level.

6.6 Definitions

- Baseline: In general, baseline will be defined as the last assessment result on or before the randomization date (and time if available).

  For safety assessment data (laboratory, vital sign): for treated patients, baseline will be defined as the last assessment result on or before the date (and time if available) of the first dose of study treatment; for patients who are randomized but not treated, baseline will be defined as the last assessment result on or before the randomization date.

  If there are multiple valid safety assessments on or prior to the first dose of study treatment:

  - For laboratory tests, the latest non missing lab value on or before the first dose date (and time if collected) will be used as the baseline in the analyses. If multiple observations exist on the latest collection date (and time if collected), then the first observation is used as baseline.

  - For ADA, the record related to the most recent assessment among those records where date (and time if collected) of the ADA assessment being performed for a specific drug is on or before to the date (and time if collected) of the first dose date of that specific drug.

  Pre-treatment AEs will be defined as AEs with an onset date prior to, but not including, the day of the first dose of study treatment.

- Study Day: There is no study day zero. In general, Study Day 1 is the date of randomization.
  - For post-randomization days, study day is calculated as:
    assessment date – randomization date + 1
  - For pre-randomization days, study day is calculated as:
    assessment date – randomization date

  For safety analysis, study day is calculated and presented based on the first dose date as defined below.

- First Dose Date: First dose date will be defined as the date of first dose of study treatment, Cycle 1 Day 1 (C1D1). Note, per protocol, the first dose date may be ≤ 5 days after the date of randomization. If the randomization date and the first dose date is not on the same day, then the date of randomization will be different from C1D1. Safety data will be calculated and presented based on First Dose Date.
- Treatment-emergent/Extended Treatment-emergent Period:

  For patients on treatment, both the treatment-emergent period and the extended treatment-emergent period will be defined as the period of time on or after the day (and time, if collected and not missing) of the first dose of study treatment.

  For patients off treatment, the treatment-emergent period is defined as the period of time on or after the day (and time, if collected and not missing) of the first dose of study treatment, until: the date of (initiation of new anticancer therapy – 1 day) or (30 days after the date of the last dose of any study treatment), whichever is earlier.

  The extended treatment-emergent period will be defined as the period of time on or after the day (and time, if collected and not missing) of the first dose of study treatment until earlier date of (initiation of new anticancer therapy – 1 day) and (100 days after the date of the last dose of any study treatment).
- Treatment-Emergent Adverse Event (TEAE):

TEAEs are AEs with an onset date (and time, if collected and not missing) within the treatment-emergent period, or extended treatment-emergent period depending on the analysis

Incomplete start and end dates for TEAEs will be imputed. Any AE will be considered as a TEAE if its status cannot be fully determined because of incomplete data.

- Last Known Alive Date: Last known alive date will be defined as the latest alive date on or before data cutoff date, captured from multiple sources. Note that if there is an event date (either an assessment or death) beyond the data cutoff date, the last known alive date will be the data cutoff date. If a patient died before or on the data cutoff date, the last known alive date will be set as the death date. The last known alive date will not be imputed, except in the case that the death date is partially missing (see details in Section 6.4)

6.7 Study Drug Exposure

For NKTR-214 and nivolumab in Arm A, the following parameters will be calculated:

- Exposure duration (days) for each drug: Date of last dose – date of first dose (C1D1) + 1

For NKTR-214 in Arm A, the following parameters will be calculated:

- Total number of infusions: Total number of infusions for which the patient received a non-zero dose across all cycles

- Cumulative dose (mg/kg): Total actual dose (mg/kg) the patient received across all cycles, defined as the sum of actual doses (mg/kg) received across all cycles
- Average dose per infusion (mg/kg): Cumulative dose (mg/kg) / Total number of infusions
- Average duration of infusion (min): Total infusion time (min) / Total number of infusions; where total infusion time = summation of (completion time of infusion – start time of infusion) for all infusions of a patient
- Actual dose intensity (mg/kg/week): [Cumulative dose (mg/kg) / (Exposure duration (in days) + 20 days)] × 7
- Expected dose intensity (mg/kg/week) = (0.006 mg/kg) / (3 weeks) = 0.002 mg/kg/week
- Relative dose intensity (%): (Actual dose intensity / Expected dose intensity) × 100

For Nivolumab in Arm A, the following parameters will be calculated:

- Total number of infusions: Total number of infusions for which the patient received a non-zero dose across all cycles
- Cumulative dose (mg): Total actual dose (mg) the patient received across all cycles, defined as the sum of actual doses (mg) received across all cycles
- Average dose per infusion (mg): Cumulative dose (mg) / Total number of infusions
- Average duration of infusion (min): Total infusion time (min) / Total number of infusions; where total infusion time = summation of (completion time of infusion – start time of infusion) for all infusions of a patient
- Actual dose intensity (mg/week): [Cumulative dose (mg) / (Exposure duration (in days) + 20 days)] × 7
- Expected dose intensity (mg/week) = (360 mg) / (3 weeks) = 120 mg/week
- Relative dose intensity (%): (Actual dose intensity / Expected dose intensity) × 100

For Sunitinib in Arm B, the following parameters will be calculated:

- Exposure duration (days): Date of last dose – date of first dose (C1D1) + 1
- Total number of doses: Total number of doses the patient received
- Average daily dose (mg/day): Sum of all sunitinib doses in mg actually received / Exposure of duration (in days)
- Expected daily dose (mg/day): 33.33 mg/day (= 50 mg × 28 days / 42 days)

  (Note that sunitinib treatment consists of 50 mg PO daily dose for 4 weeks followed by 2 weeks of a washout period)
- Relative dose intensity (%): (Average daily dose / Expected daily dose) × 100

For Cabozantinib in Arm B, the following parameters will be calculated:

- Exposure duration (days): Date of last dose – date of first dose (C1D1) + 1
- Total number of doses: Total number of doses the patient received
- Average daily dose (mg/day): Sum of all cabozantinib doses in mg actually received / Exposure of duration (in days)
- Expected daily dose (mg/day): 60 mg/day or 40 mg/day (40 mg for patients with mild to moderate hepatic impairment [Child-Pugh score A or B])
- Relative dose intensity (%): (Average daily dose / Expected daily dose) × 100

7.0 STATISTICAL ANALYSIS

7.1 Patient Disposition

A summary of patient disposition will display the number of patients who were randomized and the number that comprised each analysis population by treatment arm for the ITT I/P-risk population and ITT all-risk population. In addition, the number of patients who discontinued study drug and the number of patients who exited the study, both overall and by reason, will be presented by treatment arm for the ITT I/P-risk population and the ITT all-risk population. Descriptive statistics for time to NKTR-214 discontinuation (mean, median, min, max) will be summarized for the patients who discontinued NKTR-214.

All disposition data will be presented in a data listing.

7.2 Important Protocol Deviations

Important protocol deviations are defined as protocol deviations that fall into the following categories:

- Violated key inclusion/exclusion criteria
    - Patient did not have advanced or metastatic RCC with a clear-cell component
    - Patient received prior systemic therapy
    - Patient had active brain metastases
    - Patient did not meet inclusion/exclusion criteria for lab values (e.g., absolute neutrophil count [ANC], platelet count, hemoglobin, serum creatinine, alanine aminotransferase [ALT], aspartate aminotransferase [AST], total bilirubin, urine protein creatinine ratio [UPCR])
    - Patient had a Karnofsky Performance Status of less than 70%
    - Patient required more than 2 anti-hypertensive medications for managing hypertension
    - Patient received prior palliative radiotherapy within 2 weeks prior to randomization
- Received wrong treatment or significant incorrect dose
    - Treated differently from the randomized assignment
    - Relative dose intensity >125% for study drug (programmatically derived). Received important prohibited medication during study participation
    - Any antineoplastic therapy (i.e., chemotherapy, hormonal therapy, immunotherapy, extensive non-palliative radiation therapy, or investigational agent)
- Missing critical primary/key secondary efficacy assessment
    - The baseline disease assessment is missing in BICR data (programmatically derived)

- Stratification error
- Issues that affect patient rights
    - Patient did not sign the informed consent for the study before randomization (programmatically derived)
    - Patient developed withdraw criteria during the study but were not withdrawn (manual review)
- Other issues that may significantly impact the completeness, accuracy, and/or reliability of key study data or that may significantly affect a patient's rights, safety, or well-being

Final determination of important protocol deviations will be reviewed by the study team (including Clinical Trial Lead (CTL), the Medical Monitor, and the statistician).

The number and percentage of patients in each important protocol deviations category and sub-category will be summarized by treatment arm for the ITT I/P-risk population and ITT all-risk population. All protocol deviations will be listed.

7.3 Study Population

7.3.1 Demographics and Other Baseline Disease Characteristics

The following baseline data will be summarized and listed by treatment arms for ITT I/P-risk population and ITT all-risk population:

- Age
- Age categorization I (< 65, ≥ 65 years)
- Age categorization II (<65, ≥ 65 and < 75, ≥ 75 years)
- Age categorization III (<65, ≥ 65 and < 75, ≥ 75 and <85, ≥ 85 years)
- Sex (male, female)
- Ethnicity
- Race (White, Black, Asian, Other)
- Height
- Weight
- Geographic region (US, Rest of the world)
- Karnofsky performance status (<80, ≥80)
- IMDC prognostic score as per IRT (0, 1-2, 3-6)
- IMDC prognostic score as per EDC (0, 1-2, 3-6)
- TKI choice as per IRT (Sunitinib, Cabozatinib)

- Prior nephrectomy (Yes, No)
- Prior radiotherapy (Yes, No)
- Baseline liver mets as per BICR (Yes, No)
- LDH level (≤ 1.5 × ULN, > 1.5 × ULN)
- Hemoglobin (< LLN, ≥ LLN)
- Corrected Calcium (≤ ULN, > ULN)
- Absolute neutrophil count (ANC)
- Platelet (≤ ULN, > ULN)
- Alanine aminotransferase (ALT)
- Aspartate aminotransferase (AST)
- Alkaline phosphatase (≤ ULN, > ULN)
- PD-L1 status (≥ 1%, <1%, indeterminate or non-evaluable)
- All lesions (Investigator and BICR Assessments at Baseline): sites of disease, number of disease sites per patient
- Target lesions (Investigator and BICR Assessments at Baseline): presence of target lesions, site of target lesions, sum of reference diameter of target lesions, number of disease sites per patient
- Hepatic impairment (No [total bilirubin ≤ ULN and AST ≤ ULN], Yes [total bilirubin > ULN or AST > ULN])
- Renal impairment (No [creatinine clearance ≥ 60 mL/min], Yes [creatinine clearance < 60 mL/min])

7.3.2 Medical and Cancer History

Cancer history will be summarized and listed for the ITT I/P-risk population and the ITT all-risk population, and will include the time since initial RCC cancer diagnosis, time since metastatic/advanced RCC cancer diagnosis, history of brain metastases, histology and sarcomatoid component.

Medical history collected at screening will be mapped by the Medical Dictionary for Regulatory Activities (MedDRA) and will be summarized by system organ class (SOC) and preferred term for the ITT I/P-risk population and the ITT all-risk population. For summary tables, a patient will be counted only once per SOC and preferred term.

7.3.3 Prior Systemic Cancer/Oncology Therapies

An eligible patient should not receive any prior systemic cancer/oncology therapies for RCC. All patients who received prior systemic cancer/oncology therapies for RCC are considered to have important protocol violations.

All prior systemic cancer/oncology therapies will be listed, if applicable.

7.3.4 Surgical History and Prior Radiotherapy

The number and percent of patients who had prior cancer-related surgery will be summarized by type of surgery for each treatment arm using the ITT I/P-risk population and ITT all-risk population. Patients who have undergone prior nephrectomy will be summarized.

The number and percent of patients who had prior radiotherapy will be summarized by type of radiotherapy for each treatment arm using the ITT I/P-risk population and ITT all-risk population.

All prior radiotherapy and surgical procedures will be listed.

7.3.5 Concomitant Radiotherapy and Procedures

The number and percent of patients who had concomitant radiotherapy will be summarized by site and type of radiotherapy for each treatment arm using the ITT I/P-risk population and ITT all-risk population.

The number and percent of patients who had concomitant procedures will be summarized by each coded procedure preferred term for each treatment arm using the ITT I/P-risk population and ITT all-risk population.

All concomitant radiotherapy and procedures will be listed.

7.4 Treatments and Medications

7.4.1 Prior and Concomitant Medications

Prior and concomitant medications will be coded according to Anatomical, Therapeutic Chemical Level (ATCL) classification and preferred drug name using the World Health Organization Drug Dictionary Enhanced (WHO-DDE).

Concomitant medications are defined as medications taken on or after the date of first dose of study drug, including medications initiated prior to the date of first dose of study drug and continued during treatment, and medications initiated on or after the date of first dose of study drug and within 100 days after last dose of study drug or initiation of new anticancer therapy – 1 day, whichever is earlier.

Concomitant medications will be tabulated for the Safety I/P-risk population and the Safety all-risk population by WHO-DDE ATCL-2 classifications and preferred term. If the ATCL-2 classification is missing, the next non-missing higher level of classification will be used (Level 1). If a patient reports the same medication multiple times, then the frequency reported for that medication will be incremented by only one. As with the medication, if a patient reports multiple medications within the same ATCL-2 classification then the frequency for that ATCL-2 classification will be incremented by only one. Percentages will be calculated using the total number of patients in the Safety I/P-risk population and the Safety all-risk population.

Prior medications are defined as medications taken starting prior to the first dose of study drug. Prior medications will be summarized for the ITT I/P-risk population and ITT all-risk population using the same analytical procedures as concomitant medications. For the medications initiated prior to the date of first dose of study drug and continued during treatment, they are considered as both prior medications and concomitant medications.

Number of patients that received narrow-range CYP substrates as concomitant medication will also be summarized.

Prior and concomitant medications will be presented in a data listing.

7.4.2 Immune Modulating Medication

The list of anatomic class, therapeutic class and generic name used for the selection of immune-modulating medication (IMM) will be provided at the time of the database lock.

The percentage of patients who received immune-modulating concomitant mediations for

- Management of drug-related select adverse events (any Grade, Grade 3-5) by select AE category/subcategory (see Section 8.2 for select AEs)
- Management of IMAEs (any grade, grade 3-5) by IMAE category (see Section 8.2.2 for IMAEs) will be reported by medication class and generic term.

For each category/subcategory of drug-related select AEs (any grade, grade 3-5) and IMAEs (any grade, grade 3-5), the following will be reported:

- Total duration of IMM use (excluding overlaps), duration of high dose of corticosteroid, initial dose of corticosteroid, and tapering duration (summary statistics)

7.4.3 Subsequent Anti-Cancer Therapy

Subsequent anti-cancer systemic therapy will be summarized based on the ITT I/P-risk population and the ITT all-risk population, and tabulated using the WHO-DDE and ATCL-2 classifications and preferred term. The number and percent of patients who took at least one subsequent anti-cancer therapy will be calculated and presented by treatment group, ATCL-2 classifications, and preferred term.

7.4.4 Study Drug Exposure

Overall exposure to each study drug will be summarized for the Safety I/P-risk population and the Safety all-risk population. Overall exposure to each study drug will be summarized in terms of exposure duration, number of infusions/doses, and relative dose intensity. The following summaries will also be performed:

- Time from randomization to first dose of study therapy (0 to 5 days, > 5 to 7, > 7 to 14, > 14 to 21, > 21 to 28, > 28)
- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%
- For Arm A, the number of patients with at least one dose infusion interruption, the reason for interruption, and the number of infusion interruptions per patient
- For Arm A, the number of patients with at least one dose delay (either study drug) or dose reduction (for NKTR-214 only) and their reason(s)
- For Arm B, the number of patients with dose reduction and dose missed will be tabulated along with their reason(s)
- For NKTR-214, the time to first dose reduction and time to first dose infusion interruption will be summarized using descriptive statistics (mean, median, minimum, maximum)

All study exposure data will be presented in a data listing.

7.5 Efficacy Analysis

7.5.1 Efficacy Analysis of Primary Endpoints

The primary endpoints are ORR by BICR in the ITT I/P-risk population and the ITT all-risk population, and OS in the ITT I/P-risk population and the ITT all-risk population. The overall alpha for this study is 0.05, which is split with 0.001 to evaluate ORR in the ITT I/P-risk population and the ITT all-risk population, and 0.049 to evaluate OS in the ITT I/P-risk population and the ITT all-risk population. All alpha levels are two-sided. The test in the ITT all-risk population is gated by the one in ITT I/P-risk population for both ORR and OS. That is, ORR in the ITT all-risk population will not be tested unless ORR in the ITT I/P-risk population is statistically significant; and similarly, OS in the ITT all-risk population will not be tested unless OS in the ITT I/P-risk population is statistically significant.

In addition, a fallback approach will be used to reallocate alpha from ORR to OS. If ORR is significant in the ITT I/P-risk population and the ITT all-risk population, the alpha of 0.001 will be passed to OS. OS in the ITT I/P-risk population can then be tested at 0.05 (instead of 0.049) level. If it is significant, the alpha of 0.05 will be passed to OS in the ITT all-risk population.

7.5.1.1 Objective Response Rate

Objective response rate (ORR) is defined as the number of randomized patients who achieve a best overall response of confirmed complete response (CR) or confirmed partial response (PR) based on BICR assessments (using mRECIST 1.1 criteria) divided by the number of all randomized patients. Best Overall Response (BOR) is defined as the best response, as determined by the BICR, recorded between the date of randomization and the date of objectively documented progression per mRECIST 1.1 criteria or the date of subsequent anti-cancer therapy (including tumor-directed radiotherapy and tumor-directed surgery), whichever occurs first. For patients without documented progression or subsequent anti-cancer therapy, all available response designations will contribute to the BOR determination. Confirmation of response is required at least 4 weeks after the initial response. A Best Overall Response of SD requires a minimum of 56 days on study from randomization to the date of the first imaging assessment.

The primary analysis is based on the ORR I/P-risk Population and the ORR All-risk Population. The ORR will be compared between treatment arms using a Cochran-Mantel Haenszel (CMH) test stratified by the stratification factors for analysis (IMDC prognostic score and TKI choice) at a 0.001 alpha level. The ORR difference between the two treatment arms with its 95% and 99.9% confidence interval (CI) will be calculated using CMH methodology and adjusted by the same stratification factors. The ORR and its corresponding 95% exact CI will also be calculated by Clopper-Pearson method (Clopper and Pearson, 1934) for each arm. The number and percentage of patients in each category of BOR per BICR (complete response [CR], partial response [PR], stable disease [SD], progressive disease [PD], or not evaluable [NE]) will be presented, by treatment arm.

The ORR per investigator assessment will also be analyzed as a sensitivity analysis, following the same approach as above. The ORR concordance rate will be computed as the frequency with which the investigator's assessment and BICR's assessment agree on classification of a patient as responder versus non responder.

The ORR for patients who received sunitinib and the ORR for patients who received NKTR-214 and nivolumab will be summarized and compared using the CMH test stratified by the stratification factor IMDC prognostic score.

7.5.1.2 Overall Survival

Overall Survival (OS) is defined as the time from the date of randomization to death from any cause. Patients will be followed until their date of death, loss to follow-up, or withdrawal of consent for further follow-up for survival. Patients who are lost to follow-up will be censored at their last known alive date. Patients who do not have date of death at final database lock will be censored at their last known alive date. Patients who do not have any follow-up since randomization will be censored on the date of randomization.

The primary analysis for OS will use the ITT I/P-risk Population and the ITT All-risk Population.

Two-sided alpha levels of 0.01 and 0.0472 are allocated to the interim and final analyses, respectively, to control the overall type I error rate at 0.05 level for the primary OS analysis in the IMDC Intermediate- or Poor-risk and the IMDC All-risk Populations assuming ORR is significant in both the IMDC Intermediate- or Poor-risk and the IMDC All-risk Populations. If ORR is not statistically significant in either population, an overall alpha of 0.049 will be used for OS where 0.01 will be spent at the interim analysis and 0.0462 will be spent at the final analysis. Within each analysis, OS in the IMDC Intermediate- or Poor-risk Population will be tested first, and only when it is statistically significant, can OS in the IMDC All-risk Population be tested at the same level.

Overall survival will be compared between the treatment groups at the interim and final analyses, using the stratified logrank test. The stratification factors will be the same as those used in the analysis of ORR (IMDC prognostic score and TKI choice) as defined in Section 6.5. At the final analysis, the primary analysis of OS to claim statistical significance will be based on the weighted combination test of the stratified logrank test statistic with pre-specified weights of $\sqrt{2/3}$ and $\sqrt{1/3}$ (Lehmacher and Wassmer, 1999). The details are included in Appendix 15.1. The conventional stratified logrank test with equal weights for every patient will be conducted as a sensitivity analysis at the final analysis.

A stratified Cox proportional hazards model with treatment as the single covariate will be used to estimate the hazard ratio and corresponding 95% CI. The stratification factors will be the same as those used in the analysis of ORR (IMDC prognostic score and TKI choice) as defined in Section 6.5. The Kaplan-Meier method will be used to further summarize OS, including Kaplan-Meier curves, and medians with corresponding 95% CIs. The OS rates at fixed timepoints (e.g. 12 months, 24 months, depending on minimum follow-up) will also be estimated along with their corresponding 95% CIs.

A sensitivity analysis for OS using an unstratified logrank test will also be performed. The hazard ratio associated with treatment will be presented along with the corresponding 95% CI.

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

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

The OS for patients who received sunitinib and the OS for patients who received NKTR-214 and nivolumab will be summarized and compared using the Kaplan-Meier method. The median and 95% CI will be presented for each treatment group, and the stratified logrank test will be performed. The stratification factor IMDC prognostic score will be used.

7.5.2 Efficacy Analysis of Key Secondary Endpoint

The key secondary endpoint of PFS by BICR will be tested hierarchically given the statistical significance of OS. Statistical test will be two-sided with a type I error rate of 0.05 if both ORR by BICR and OS are significant in both the IMDC Intermediate- or Poor-risk and the IMDC All-risk Populations; the type I error rate will be 0.049 if OS (but not ORR by BICR) is significant in both the IMDC Intermediate- or Poor-risk and the IMDC All-risk Populations.

The PFS interim and final analysis will be conducted at the same time as the OS interim and final analyses. Two-sided alpha levels of 0.03 and 0.0335 are allocated to the PFS in the interim and final analyses, respectively, to control the overall type I error rate at two-sided 0.05. If the overall type I error rate is to be controlled at two-sided 0.049, two-sided alpha levels of 0.03 and 0.0323 will be allocated to the PFS interim and final analyses. Details are provided in Appendix 15.2.

7.5.2.1 Progression-Free Survival

The definition for PFS accounts for subsequent anti-cancer therapy by censoring at the last evaluable tumor assessment on or prior to the date of subsequent anti-cancer therapy. Clinical deterioration in the absence of unequivocal evidence of progression per mRECIST 1.1 is not considered progression for purposes of determining PFS by BICR. PFS rates by BICR at fixed time points (e.g. 6 months, depending on the minimum follow-up) are defined as the probability that a patient has not progressed and is alive at time T following randomization.

The definition of PFS (PFS truncated at subsequent anti-cancer therapy) is defined as the time between the date of randomization and the date of first documented tumor progression based on BICR assessments per mRECIST v1.1, or death due to any cause, whichever occurs first, prior to subsequent anti-cancer therapy.

Patients who die without a reported progression will be considered to have progressed on the date of their death. The following censoring rules will be applied for the primary definition of PFS:

- Patients who did not have any on study tumor assessments and did not die will be censored on their date of randomization.
- Patients who did not receive subsequent anti-cancer therapy, did not progress and did not die will be censored on the date of their last evaluable tumor assessment.
- Patients who received subsequent anti-cancer therapy prior to documented progression or death will be censored at the date of the last evaluable tumor assessment conducted on or prior to the date of initiation of the subsequent anti-cancer therapy.
- Patients who received subsequent anti-cancer therapy without documented progression or death will be censored at the date of the last evaluable tumor assessment conducted on or prior to the initiation of the subsequent anti-cancer therapy.

The potential of a delayed effect in immunotherapy interventions may cause a late separation in the PFS Kaplan-Meier curves and non-proportional hazards may be observed, as was seen for PFS in the Checkmate-214 study (Motzer 2018b). Thus, the primary analysis of PFS will account for the expected delayed effect and compare the two treatment arms via a stratified weighted logrank test (Harrington and Fleming 1982). The two-sided stratified weighted logrank p-value will be reported using $G\ (\rho = 0, \gamma = 1)$ weights. The stratification factors will be the same as those used in the analysis of ORR (IMDC prognostic score and TKI choice) as defined in Section 6.5. At the final analysis, the primary analysis of PFS to claim statistical significance will be based on the weighted combination test of the stratified weighted logrank test statistic with pre-specified weights of $\sqrt{2/3}$ and $\sqrt{1/3}$ (Lehmacher and Wassmer, 1999). The details are presented in the Appendix.

The estimate of the PFS hazard ratio between treatment groups will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. The stratification factors will be the same as those used in the analysis of ORR (IMDC prognostic score and TKI choice) as defined in Section 6.5. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio will also be presented. The Kaplan-Meier method will be used to further summarize PFS, including Kaplan-Meier curves, and medians with corresponding 95% CIs.

For the primary analysis of PFS, the censoring methods described in Table 3 will be used.

The conventional stratified logrank test with equal weights for every patient will be conducted as a sensitivity analysis.

The source of PFS event (progression or death) will be summarized by treatment group.

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

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdraw consent, never treated)
- No baseline tumor assessment
- No on-study tumor assessment and no death
- Received subsequent anticancer therapy

PFS per investigator assessment will also be analyzed as a sensitivity analysis. The PFS by investigator will be defined as in Table 3, and it will be analyzed in the same approach following the primary definition of PFS by BICR.

Table 3: PFS Primary Analysis: Date of Progression or Censoring

Situation	Date of Disease Progression or Censoring	Outcome
No baseline tumor assessments*	Date of randomization	Censored
No on study tumor assessments and no death*	Date of randomization	Censored
Subsequent anti-cancer therapy started without death or progression per mRECIST v1.1 reported prior or on the same day	Date of last evaluable tumor assessment per mRECIST v1.1 showing no evidence of disease progression prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored
No progression and no death, and no subsequent anti-cancer therapy started	Date of last evaluable tumor assessment per mRECIST v1.1 showing no evidence of disease progression	Censored
Documented progression per mRECIST v1.1 and no subsequent anti-cancer started before	Date of the first documented progression per mRECIST v1.1 (excludes clinical progression)	Progressed
Death without progression per mRECIST v1.1 and no subsequent anti-cancer therapy started before	Date of death	Progressed

* Tumor assessments and death if any, occurring after start of subsequent anti-cancer therapy are not considered.

A sensitivity analysis of PFS accounting for two or more missing tumor assessments prior to PFS event (death or documented progression) will be conducted. A patient is considered to have two or more missing tumor assessments if the elapsed time between the PFS event and the last assessment prior to the event is greater than the time interval of two scheduled scans + 14 days (e.g., 18 weeks + 14 days if two missing tumor assessments happened before Week 54). Patients who had PFS events immediately after two or more missed tumor assessments will be censored at the date of last tumor assessment prior to the missed visits.

7.5.3 Current Status of PFS and OS Follow-Up

The extent of follow-up for survival, defined as the time between randomization date and last known date alive (for patients who are alive) or death date (for patients who died), will be summarized descriptively (median, minimum, maximum) in months for all randomized patients.

The currentness of follow-up for survival, defined as the time between last OS contact (i.e., last known alive date or death date) and clinical data cutoff date, will be summarized in months for

all randomized patients. Patients who died and patients with last known alive date on or after data cutoff date will have zero value for currentness of follow-up.

The currentness of follow-up for PFS, defined as time from last evaluable tumor assessment to cutoff date will be summarized by treatment group and overall for all randomized patients. Patients who have a PFS event will be considered as current for this analysis.

The currentness of follow-up will be categorized into the following categories: 0 month, 0-3 months, 3-6 months, 6-9 months, 9-12 months, and 12 or more months.

In addition, time to treatment discontinuation will be summarized and presented by treatment group using a Kaplan-Meier curve whereby the last dose date will be the event date for those patients who are off study therapy. Median duration of study therapy and associated 95% CI will be provided. Patients who are still on study therapy will be censored on their last dose date. Time to NKTR-214 discontinuation will be summarized in the same way as well.

7.5.4 Efficacy Analysis of Other Endpoints

7.5.4.1 Progression-Free Survival 2

PFS on next-line therapy (PFS2) is defined as the time from randomization to documented progression, per investigator assessment, after the next line of therapy or to death from any cause, whichever occurs first. Patients who were alive and without progression after the next line of therapy will be censored at last known alive date. The following censoring rules will be applied for PFS2:

For a patient who did not receive subsequent anti-cancer therapy:

- If the patient died, the patient is considered as an event on the date of death
- Else, the patient is censored at his/her last known alive date

For a patient who received subsequent anti-cancer therapy:

- If the patient had disease progression after the start of subsequent anti-cancer therapy, the patient is considered as an event on the date of PD
- Else if the patient died or started a second subsequent anti-cancer therapy, the patient is considered as an event at the death date or at the start date of second subsequent therapy whichever is the earliest.
- Else, the patient is censored at his/her last known alive date

PFS2 will be analyzed as follows:

- The PFS2 will be compared between treatment arms using a two-sided logrank test, stratified by the stratification factors for analysis (IMDC prognostic score and TKI choice) as defined in Section 6.5. The HR for PFS2 with its corresponding two-sided 95% CI will also be provided, estimated via a stratified Cox model with the stratifications factors specified above and treatment arm as the only covariate in the model.
- The Kaplan-Meier method will be used to further summarize PFS2, including Kaplan-Meier curves, and medians with corresponding 95% CIs.

7.5.4.2 Concordance Between BICR and Investigator Assessment of Progression

For the purpose of assessing concordance between the BICR and investigator tumor assessments, progression status will be categorized as documented progression, death or censored. A cross tabulation between the BICR and the investigator progression status will be presented, by treatment group. The primary analysis of PFS (ITT I/P-risk population and ITT all-risk population) will be used for this assessment.

A by patient lising of BICR and investigator PFS status and the time between progression dates according to the BICR and the investigator will be provided.

7.5.4.3 Duration of Response

Analysis of duration of response (DOR) will be performed for responders (confirmed CR or PR) only. Duration of Response (DOR) is defined as the time between the date of first confirmed documented response (CR or PR) to the date of the first documented tumor progression as determined by the BICR per mRECIST 1.1, or death due to any cause, whichever occurs first. Patients who start subsequent anti-cancer therapy without a prior reported progression will be censored at the last evaluable tumor assessments prior to initiation of the subsequent anticancer therapy. Patients who die without a reported prior progression will be considered to have progressed on the date of their death. For patients who neither progress nor die, DOR will be censored on the date of their last evaluable tumor assessment. The DOR for each treatment group will be estimated using the Kaplan-Meier method. Median values of DOR with corresponding 95% CI for each treatment group will be computed based on a log-log transformation method.

7.5.4.4 Time to Response

A descriptive analysis of time to response (TTR) will be performed for responders (confirmed CR or PR) only. Time to response is defined as the time from the date of randomization to date of first documented CR or PR per mRECIST 1.1 as assessed by BICR. TTR will be summarized using descriptive statistics only (median, min, max). TTR per investigator's assessment will be summarized similarly.

7.5.4.5 Clinical Benefit Rate

The CBR per both BICR and investigator's assessment, and its corresponding 95% exact CI will be calculated by the Clopper and Pearson method for each treatment arm.

7.5.5 Subgroup Analyses

The following subgroup analyses will be performed for ORR by BICR, OS, and PFS by BICR.

The benefit of NKTR-214 in combination with nivolumab compared to TKI will be evaluated by a unstratified rate difference for ORR by BICR or unstratified hazard ratio for OS and PFS by BICR with its 95% CI based on a Cox regression model for each subgroup. The hazard ratios and their 95% CI will be displayed in a forest plot. Subgroup analyses may not be performed if the sample size is too small to provide accurate estimates. The following subgroup analyses (except age, race, region, and sex) will be conducted if the number of patients in each subgroup is > 10 for each treatment group.

- IMDC prognostic score (0 [favorable risk], 1-2 [intermediate risk], 3-6 [poor risk])
- Baseline PD-L1 status (≥ 1%, <1%, indeterminant or non-evaluable)
- Age categorization (< 65, ≥ 65 years)
- Sex (male, female)
- Race (White, Black, Asian, Other)
- Geographic region (US, Rest of the world)
- Baseline Karnofsky performance status (<80, ≥80)
- Sarcomatoid histology (Yes vs No)
- Prior nephrectomy (Yes, No)
- Baseline bone mets as per BICR (Yes, No)

The following subgroups may be explored:

- TKI choice (sunitinib, cabozantinib)
- Baseline PD-L1 status (≥ 5%, <5%, indeterminant or non-evaluable; ≥ 10%, < 10%, indeterminant or non-evaluable)
- Time from initial RCC diagnosis to randomization (≤ 1 year, > 1 year)
- Patients with history of brain metastases (Yes, No)
- Prior radiotherapy (Yes, No)
- Baseline LDH level (≤ 1.5 x ULN, > 1.5 x ULN)
- Baseline Hemoglobin (<LLN, ≥LLN)

- Baseline Corrected Calcium (≤ ULN, > ULN)
- Baseline Absolute Neutrophil Count (≤ ULN, > ULN)
- Baseline Platelet Count (≤ ULN, > ULN)
- Baseline Alkaline phosphatase (≤ ULN, > ULN)
- Tumor burden at baseline per BICR and per investigator (≤ 1st Tertile, > 1st Tertile to ≤ 2nd Tertile, > 2nd Tertile).
- Baseline liver mets as per BICR (Yes, No)
- Baseline lung mets as per BICR (Yes, No)
- Baseline lymph node mets as per BICR (Yes, No)

7.6 Health-Related Quality of Life (HRQoL)

Patients will be asked to complete the NCCN/FACT Symptom Index for Kidney Cancer (FKSI-19), 3-level version of the EuroQol Group's EQ-5D (EQ-5D-3L), and selected items from the Patient-Reported Outcomes version of the Common Term Criteria for Adverse Events (PRO-CTCAE).

Patients will complete the patient-reported outcome measures prior to any assessments at each clinic visit before randomization, prior to the first dose on Day 1 and 22 of every cycle through Cycle 4, and then on Day 1 of all subsequent cycles. In addition, the patient-reported outcome measures will be completed at designated time points during the follow-up period. The primary time points of interest will be Week 12, Week 24, and Week 48.

7.6.1 FKSI-19

The National Comprehensive Cancer Network/Functional Assessment of Cancer Therapy–Kidney Symptom Index-19 (FKSI-19) is a 19-item self-reported questionnaire with a total score and four subscales that assess disease-related symptoms (DRS), treatment side effects (TSE), and function/well-being associated with advanced kidney cancer. It captures symptom severity and interference in activity and general health perceptions. The FKSI-19 comprises four subscales (DRS-Physical, DRS-Emotional, TSE, and Function/Well-Being). Each of the items is scored on a five-point scale from zero (not at all) to four (very much). The FKSI-19 uses a recall period of "the past 7 days", and higher FKSI-19 scores indicate improvement (Rothrock, 2013).

If there are missing items, subscale scores can be prorated. This can be done using the following formula:

> Prorated subscale score = [Sum of item scores] x [N of items in subscale] / [N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the items were answered in that subscale; otherwise the subscale is considered as not

having a valid score. The total score is then calculated as the sum of the prorated subscale scores. The total score should only be calculated if ALL of the component subscales have valid scores.

The following descriptive analyses will be conducted by treatment group in the ITT I/P-risk population and the ITT all-risk population:

- FKSI-19 questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number (ie, number of patients on treatment or in follow up), will be calculated and summarized for each assessment time point.
- For the total score and subscales, separately:
  - Mean score and mean change from baseline at each assessment time point will be summarized using descriptive statistics (N, mean, SD, median, 25th and 75th percentiles, minimum, maximum).
  - A plot summarizing the mean change from baseline will be presented.
  - Changes from baseline will be compared between treatment groups using a linear mixed model with repeated measures (MMRM) with treatment, time, treatment and time interaction, baseline score, and stratification variables as covariates. An unstructured covariance matrix will be used unless the model fails to converge. Treatment effect on the change from baseline will be compared by the difference in the least-square means and corresponding 95% CI. The MMRM assumes that the missing data mechanism is ignorable (i.e., missing at random).

7.6.2 EQ5D3L

The EQ-5D-3L is a standardized instrument used to measure self-reports of health status and functioning. The instrument's descriptive system consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 3 levels, reflecting "no health problems," "moderate health problems," and "extreme health problems." A dimension for which there are no problems is said to be at level 1, while a dimension for which there are extreme problems is said to be at level 3. Thus, the vectors 11111 and 33333 represent the best health state and the worst health state, respectively, as described by the EQ-5D-3L. Altogether, the instrument describes $3^5 = 243$ health states (EuroQol, 1990).

Empirically-derived weights can be applied to an individual's responses to the EQ-5D-3L descriptive system to generate an index measuring the value to society of his or her current health. Such preference-weighting systems have been developed for the UK, US, Spain, Germany, and numerous other populations. Utility index values range from a 1 (full health) to 0 (dead) with negative values indicating a state considered worse than being dead. In addition, the EQ-5D-3L includes a visual analog scale (VAS) that allows respondents to rate their own current health on a 101-point scale ranging from "best imaginable" to "worst imaginable" health. The EQ-5D-3L uses a recall period of "today."

EQ-5D Index values will be computed using a scoring algorithm based on the US TTO (Time Trade-off) value set. The following descriptive analyses will be conducted:

- EQ-5D questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number (ie, number of patients on treatment or in follow up), will be calculated and summarized for each assessment time point.
- A by-patient listing of EQ-5D with the problem level for each of the 5 dimensions (mobility, selfcare, usual activities, pain/discomfort and anxiety/depression), health state (5 dimensions digits combined in a 5-digit number), EQ-5D Index, and EQ-VAS will be provided.
- Proportion of patients for each level of problem for the 5 EQ-5D dimensions will be summarized at each assessment timepoint. Percentages will be based on number patients assessed at assessment time point.
- For the EQ-5D Index and visual analog scale (VAS), separately:
  - Mean score and mean change from baseline at each assessment time point will be summarized using descriptive statistics (N, mean, SD, median, 25th and 75th percentiles, minimum, maximum).
  - A plot summarizing the mean change from baseline will be presented.

7.6.3 PRO-CTCAE

The PRO-CTCAE is a patient-reported outcome measure developed to evaluate symptomatic toxicity in patients participating in oncology clinical trials. It was designed to be used as a companion to the Common Terminology Criteria for Adverse Events (CTCAE), the standard lexicon for adverse event reporting in cancer trials. PRO-CTCAE includes an item library of 124 items representing 78 symptomatic toxicities drawn from the CTCAE. PRO-CTCAE provides a systematic yet flexible tool for descriptive reporting of symptomatic treatment side effects in cancer clinical trials. Selected items from the PRO-CTCAE will be administered. The PRO-CTCAE uses a recall period of "the last 7 days."

To report the PRO-CTCAE data, the tabulation of the maximum score post-baseline per patient per item will be used, and it will be presented by treatment group. A by-patient listing of PRO-CTCAE will also be provided.

7.7 Healthcare Resource Utilization

Healthcare resource utilization (HCRU) information is collected up to safety follow up visit 1 (100 days after last dose of study treatment), and will be summarized by treatment group. The HCRU summary includes:

- Number of hospitalizations per patient (0, 1, 2, > 2) for overall and by hospitalization ward (general ward, intensive care ward, rehabitation ward, hospice ward, other ward)

- Number of hospitalizations per person-year
    - Numerator is number of hospitalizations from all patients in the analysis population.
    - Denominator is the sum of {min[(last dose date + 100 days), last known alive date] – first dose date +1} across all patients in the analysis population /365.25.
- Number of days in each specific hospitalization ward per person-year. Specific hospitalization ward includes general ward, intensive care ward, rehabitation ward, hospice ward, other ward, and overall.
    - Numerator is sum of number of days in each specific hospitalization ward from all patients in the analysis population.
    - Denominator is the sum of {min[(last dose date + 100 days), last known alive date] – first dose date +1} across all patients in the analysis population /365.25.

A by-patient listing of HCRU data will also be provided.

7.8 Immunogenicity

7.8.1 Immunogenicity Measurements and Endpoints

The immunogenicity analysis will be performed separately for anti-NKTR-214/anti-PEG, anti-IL-2 and anti-nivolumab antibodies. The data will be summarized for anti-NKTR-214/anti-PEG, anti-IL-2 and anti-nivolumab antibodies and will be reported by both sample and patient status.

For sample status, baseline samples will be classified as:

- baseline ADA-positive, if ADA is detected in the immunogenicity sample collected before initiation of the treatment and
- baseline ADA-negative, if ADA is not detected in the immunogenicity sample collected before initiation of the treatment.

ADA samples will be classified as positive if after initiation of the treatment, the ADA is detected (positive seroconversion) in a sample from a patient who was baseline ADA-negative, or if there is at least 4-fold or greater (≥) increase in ADA titer in comparison to the baseline positive titer. Samples confirmed to be positive for anti-NKTR-214/anti-PEG, anti-IL-2 antibodies will be tested for neutralizing activity for IL-2 using a validated cell-based assay. Anti-nivolumab antibodies positive samples will also be tested for neutralizing activity for nivolumab using a validated cell-based assay. ADA-positive samples with neutralizing activity will be classified as neutralizing positive.

For patient status, immunogenicity will be reported by summarizing the number and percentage of ADA-positive patients and ADA-negative patients at baseline and after drug administration with positive status of anti-NKTR-214, anti-PEG, anti-IL-2, and anti-nivolumab ADA. In the

immunogenicity evaluable population, the patients will be classified as treatment-induced ADA-positive if they were ADA-negative at baseline and became ADA-positive at any time after starting the treatment, or if they were ADA-positive at baseline and had a post-baseline ADA-positive titer that was at least 4-fold or greater (≥) than baseline positive titer (also known as Treatment-boosted). The patients with an ADA-positive sample at 2 or more consecutive time points, where the first and last ADA-positive samples are at least 16 weeks apart will be classified as persistent positive. Patients that are not persistent positive but have an ADA-positive sample at the last sampling time point will be classified as not persistent last-sample positive. Non persistent positive patients with an ADA-negative sample at the last sampling time point will be classified as transiently positive or other positive. Patients with at least 1 ADA-positive sample with neutralizing activity will be classified as neutralizing positive. All other ADA-evaluable patients are considered as treatment-emergent ADA negative.

All ADA-positive patients will be included in the ADA titer kinetics analysis. Summary statistics of patient-level ADA titers using the maximum titer value within an ADA-positive patient will be presented for baseline ADA-negative patients and baseline ADA-positive patients. The median, interquartile range, and range of the maximum titers will be reported. For ADA-positive patients with baseline ADA-positive sample, the median and interquartile range of the fold increase from baseline in titer (ratio of maximum post-baseline titer to baseline titer) will also be reported. For sample-level ADA titers, boxplots of ADA titers at each assessment timepoint will be provided, as appropriate, to demonstrate whether the ADA levels tend to change over time during the treatment, along with ADA incidence at each assessment timepoint.

Analyses of ADA kinetics, including the onset and duration of ADA, will include only those ADA-positive patients that were ADA-negative at baseline. ADA-positive patients with baseline ADA-positive sample and an increased titer post-baseline will be excluded since this type of immune response differs mechanistically.

Onset of ADA refers to the time period between the first dose of the treatment and the first instance of ADA detection. The median, interquartile range, and range of time to ADA development will be provided. The median and range of the number of treatment doses to before and after the first detection of ADA as well as the total number of doses received will also be provided.

Duration of ADA refers to the longevity of ADA and is defined as the time from the first on-study ADA-positive date to the first ADA-negative date after the last on-study ADA-positive date. Duration of ADA will be estimated by the Kaplan-Meier method. Patients who do not have an ADA-negative date after the last on-study ADA-positive date will be censored on the date of their last ADA-evaluable assessment. The median together with the corresponding 95% confidence interval (using log-log transformation method) and range (minimum, maximum) will be provided.

7.8.2 Evaluation of ADA

The following analyses will be conducted for anti-NKTR-214/anti-PEG, anti-IL-2 and anti-nivolumab, separately.

- Safety: All patients with ADA as mentioned above will be assessed for potential infusion related reactions under NKTR 214 safety events and will be compared against ADA negative patients, where infusion related reaction events that start after study drug administration on the same day or next day will be included.
- Efficacy: Objective response rate will be presented by ADA status.
- Concentration: Pre-dose NKTR-214 RC and pre-dose NKTR-214 AC concentrations will be summarized by ADA status.

7.8.3 Definitions

There are two sets of definitions: one for categorizing individual samples (Table 4) and another for categorizing patient responses (Table 5).

Table 4: ADA Status: Individual Samples

ADA Status	Definition
Baseline negative	ADA is not detected in the last sample before initiation of treatment
Baseline positive	ADA is detected in the last sample before initiation of treatment
Anti-NKTR-214 ADA-positive sample	After initiation of treatment, (1) an ADA detected (positive seroconversion) sample in a patient for whom anti-NKTR-214 ADA is not detected at baseline, *or* (2) an ADA detected sample with anti-NKTR-214 titer to be at least 4-fold or greater (≥) than baseline positive titer
Anti-PEG ADA-positive sample	Sample with a positive result in PEG-immunocompetition assay.
Anti-IL-2 ADA-positive sample	After initiation of treatment, (1) an ADA detected (positive seroconversion) sample in a patient for whom anti-IL-2 ADA is not detected at baseline, *or* (2) an ADA detected sample with anti-IL-2 titer to be at least 4-fold or greater (≥) than baseline positive titer
Anti-nivolumab ADA-positive sample	After initiation of treatment, (1) an ADA detected (positive seroconversion) sample in a patient for whom anti-Nivolumab ADA is not detected at baseline, *or* (2) an ADA detected sample with anti-nivolumab titer to be at least 4-fold or greater (≥) than baseline positive titer
ADA-negative	After initiation of treatment, assay result is not a positive sample relative to baseline, i.e., increase from baseline is <4-fold

Table 5: ADA Response Categories: Patient Level

ADA status	Description
Baseline ADA-positive	A patient with baseline ADA-positive sample
ADA-negative patient	Patient with no ADA-positive sample after the initiation of treatment. **Note:** due to the definition of an ADA positive sample for a patient testing positive at baseline (see sample status table, above), it is possible (but highly unlikely) for a patient with a stable anti-drug titer throughout the study to be classified as 'ADA-negative'
ADA-positive patient (Treatment-emergent positive)	Patient with at least 1 ADA positive-sample (relative to baseline) at any time after initiation of treatment
Persistent positive	ADA-positive sample at 2 or more consecutive time points, where the first and last ADA-positive samples are **at least 16 weeks** apart (endogenous human IgG1, IgG2, and IgG4 have approximate half-lives in the range of 21-25 days, and 5 half-lives are approximately equal to 16 weeks)
Not PP-last sample positive	Not persistent positive with ADA-positive sample at the last sampling time point
Transiently positive or other positive	Patient not persistently positive but has 1 or more ADA positive samples with the last sample being ADA-negative.
Neutralizing positive	At least 1 ADA-positive sample with neutralizing antibodies detected

Abbreviation: PP = persistent positive

8.0 SAFETY ANALYSIS

The safety data will include AEs, SAEs, clinical laboratory tests, and vital signs. Summaries will use the Safety all-risk population and will be presented separately for patients who received NKTR-214 and nivolumab or who received TKI (sunitinib or cabozantinib). Summaries using the Safety I/P-risk population may be presented. All safety data will be presented in data listings. Summaries for patients who received sunitinib and patients who received cabozantinib may be performed separately on some key safety analyses.

8.1 Adverse Events

Adverse events will be coded by system organ class (SOC) and preferred term (PT) using MedDRA. Adverse event severity will be based on NCI CTCAE Grade (version 5.0).

All AEs will be coded by MedDRA version 21.1. The severity of AE will be determined based on the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Grade (version 5.0). An AE is considered as related to treatment if it is related to any study drug; an AE with a missing relationship will be counted as related to study drug. An AE is considered as leading to treatment discontinuation if it is leading to discontinuation of any study drug. An AE is considered as leading to treatment delay, reduction or interruption if reported as leading to the dose delay, reduction or interruption of any study drug. An AE is considered as leading to death if the severity grade is 5 or the outcome is fatal. A patient reporting the same AE multiple times will be counted only once within each SOC and PT under the highest severity and closest relationship.

The following adverse events summaries will be provided by treatment arm:

- Overall summary of TEAEs
- TEAEs presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- Serious TEAEs presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- Drug-related TEAEs presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- NKTR-214-related TEAEs presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- TEAEs leading to any study drug discontinuation presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- TEAEs leading to infusion interruption of any study drug presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- TEAEs leading to dose reduction of any study drug presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT

- TEAEs leading to discontinuation of NKTR-214 presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- TEAEs leading to infusion interruption of NKTR-214 presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- TEAEs leading to dose reduction of NKTR-214 presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- Overall summary of TEAEs that required immune modulating medication by worst grade (any grade, grade 3-4, grade 5) by SOC and PT

Additionally, TEAEs will also be summarized in tables as follows:

- Grade 3 or above TEAE by PT
- Grade 3 or above drug-related TEAE by PT
- Grade 3 or above NKTR-214-related TEAE by PT
- Serious drug-related TEAE by PT
- Serious NKTR-214-related TEAE by PT
- Drug-related TEAE leading to discontinuation of any study drug
- NKTR-214-related TEAE leading to discontinuation of NKTR-214
- Drug-related TEAE leading to dose reduction of any study drug
- NKTR-214-related TEAE leading to dose reduction of NKTR-214
- Drug-related TEAE leading to death by PT
- NKTR-214-related TEAE leading to death by PT

The following analysis will be repeated using an extended treatment-emergent period.

- Drug-related TEAE presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
- NKTR-214-related TEAE presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT

All AEs will be presented in listing as follows:

- SAEs
- Grade 3 or above TEAEs
- TEAE leading to any study drug dose delay
- TEAE leading to any study drug dose reduction
- TEAE leading to any study drug infusion interruption

- TEAE leading to any study drug discontinuation
- TEAEs leading to death

Overall summary of TEAEs and TEAEs by worst grade (any grade, grade 3-4, grade 5) presented by SOC/PT and treatment group will be provided for the following subgroups:

- Gender (Male, Female)
- Race (White, Black, Asian, Other)
- Age categorization (< 65, ≥ 65 years)
- Region (US, Rest of the world)
- Hepatic impairment at baseline (No [total bilirubin ≤ ULN and AST ≤ ULN], Yes [total bilirubin > ULN or AST > ULN])
- Renal impairment at baseline (No [creatinine clearance ≥ 60 mL/min], Yes [creatinine clearance < 60 mL/min])
- TKI Choice (sunitinib, cabozantinib)
- IMDC score (0 [favorable risk], 1-2 [intermediate risk], 3-6 [poor risk])

The subgroup analysis will be conducted using treatment-emergent period only.

8.2 Adverse Events of Special Interest for Checkpoint Inhibitors

8.2.1 Select Adverse Events *(to Support EU MAA)*

The select Adverse Events (select AEs, also known as select immune-related AEs) consist of a list of preferred terms grouped by specific category (e.g., pulmonary events, gastrointestinal events categories, etc.). Select AEs are identified based on the following 4 guiding principles:

- AEs that may differ in type, frequency, or severity from AEs caused by non-immunotherapies
- AEs that may require immunosuppression (e.g., corticosteroids) as part of their management
- AEs whose early recognition and management may mitigate severe toxicity
- AEs for which multiple event terms may be used to describe a single type of AE, thereby necessitating the pooling of terms for full characterization.

Based on these guiding principles and taking into account the types of AEs already observed across studies of nivolumab monotherapy, endocrinopathies, diarrhea/colitis, hepatitis, pneumonitis, interstitial nephritis, and rash are currently considered to be select AEs. Multiple event terms that may describe each of these are grouped into endocrine, gastrointestinal (GI), hepatic, pulmonary, renal, and skin select AE categories, respectively.

Hypersensitivity/infusion reactions are analyzed along with the select AE categories because multiple event terms may be used to describe such events and pooling of terms was therefore necessary for full characterization. Hypersensitivity/infusion reactions do not otherwise meet criteria to be considered select AEs.

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

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

Further details on the definitions of select adverse event, time-to onset and time-to resolution are described in Appendix 15.4.

Unless otherwise specified, analyses will be performed by select AE category. Analyses will also be repeated by subcategory of endocrine events.

8.2.1.1 Incidence of Select AE

Select AEs (also known as select immune related AEs) will be summarized using the treatment-emergent period by treatment arm for each category/subcategory.

- Overall summaries of any select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by category or subcategory/PT
- Overall summaries of any drug-related select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by category or subcategory/PT
- Overall summaries of any serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by category or subcategory/PT
- Overall summaries of drug-related serious select AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by category or subcategory/PT
- Overall summaries of any select AEs leading to any study drug discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by category or subcategory/PT
- Overall summaries of drug-related select AEs leading to any study drug discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by category or subcategory /PT
- Summary of frequency of unique select AEs by category

A by-patient listing of select AE will be provided.

8.2.1.2 Time-to Onset of Drug-Related Select AE

Time-to onset of drug-related select AEs (any grade, grade 3-5) will be summarized for each category/subcategory by treatment arm. The summary is descriptive with median, min and max.

Time-to onset analyses are restricted to treated patients who experienced at least one drug-related select AE in the category/subcategory.

Additional details regarding the time-to onset definition are described in Appendix Section 15.4.1.

8.2.1.3 Time-to Resolution of Drug-Related Select AE

Time-to resolution of the following specific events will be summarized separately for each category/subcategory by treatment arm.

- Time-to resolution of drug-related select AE (any grade, grade 3-5)
- Time-to resolution of drug-related select AE (any grade, grade 3-5) where immune modulating medication was initiated

Time-to resolution analyses are restricted to treated patients who experienced the specific events. Time-to resolution where immune-modulating medication was initiated analyses are restricted to treated patients who experienced the specific events and who received immune-modulating medication during the longest select AE.

The following summary statistics will be reported: percentage of patients with resolution of the longest select AE, median time-to resolution along with 95% CI (derived from Kaplan-Meier estimation) and ranges.

Additional details regarding the time-to resolution definition are described in Appendix Section 15.4.2.

8.2.2 Immune Mediated Adverse Events (IMAEs) *(to Support US BLA)*

Immune-mediated AEs (IMAE) are AEs consistent with an immune-mediated mechanism or immune-mediated component for which non-inflammatory etiologies (eg, infection or tumor progression) have been ruled out. IMAEs can include events with an alternate etiology which were exacerbated by the induction of autoimmunity. To be specific, IMAEs are events (or groups of PTs describing specific events) that include pneumonitis, diarrhea/colitis, hepatitis, nephritis/renal dysfunction, rash, endocrine (adrenal insufficiency, hypothyroidism/thyroiditis, hypothyroidism, thyroiditis, hyperthyroidism, diabetes mellitus and hypophysitis), and other specific events, considered as potential immune-mediated events by investigator that meet the definition summarized below:

- those occurring within extended treatment-emergent period,

- regardless of causality,
- treated with immune-modulating medication (of note, endocrine AEs such as adrenal insufficiency, hypothyroidism/thyroiditis, hypothyroidism, thyroiditis, hyperthyroidism, diabetes mellitus, and hypophysitis are considered IMAEs regardless of immune-modulating medication use, since endocrine drug reactions are often managed without immune-modulating medication),
- with no clear alternate etiology based on investigator assessment, or with an immune-mediated component

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

IMAEs will be summarized using the extended treatment-emergent period for each immune-mediated category for Arm A only:

- Overall summary of non-endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) where immune-modulating medication was initiated presented by category / PT.
- Overall summary of endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by category / PT
- Overall summary of non-endocrine IMAEs leading to study drug discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by category / PT
- Overall summary of endocrine IMAEs leading to study drug discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by category / PT
- Overall summary of non-endocrine IMAEs leading to study drug dose delay or reduction by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by category / PT
- Overall summary of non-endocrine IMAEs leading to study drug dose delay by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by category / PT
- Overall summary of non-endocrine IMAEs leading to study drug dose reduction by worst CTC grade (any grade, grade 3-4, grade 5) where immune modulating medication was initiated presented by category / PT
- Overall summary of endocrine IMAEs leading to study drug dose delay or reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by category / PT
- Overall summary of endocrine IMAEs leading to study drug dose delay by worst CTC grade (any grade, grade 3-4, grade 5) presented by category / PT

- Overall summary of endocrine IMAEs leading to study drug dose reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by category / PT
- Summaries of time-to onset and time-to resolution of non-endocrine IMAEs where immune modulating medication was initiated presented by category
- Summaries of time-to onset and time-to resolution of endocrine IMAEs presented by category.

A by-patient listing of IMAEs will be provided. By-patient listings of time-to resolution for longest IMAEs cluster (any grade and grade 3-5 in separate summaries) will also be provided.

8.2.3 Other Events of Special Interest

Other events of special interest (OEOSI, also known as other immue-related adverse events of special interest) consist of a list of preferred terms grouped by specific category (e.g. Myositis Event, Myocarditis Event, Demyelination Event, Guillain-Barre Syndrome, Pancreatitis Event, Uveitis Event, Encephalitis Event, Myasthenic Syndrome, Rhabdomyolysis Event, Graft Versus Host Disease).

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

OEOSI will be summarized by treatment arm for each category.

The following analyses will be conducted using the extended treatment-emergent period:

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

A by-patient listing of OEOSI will be provided.

8.3 Adverse Events for NKTR-214

8.3.1 Adverse Event of Special Interest for NKTR-214

Adverse events of special interest for NKTR-214 consists of ischemic cerebrovascular events (ICE). All analyses will be conducted in the Safety I/P-risk population and the Safety all-risk population by treatment arm using the treatment-emergent period. Search list used to identify ICE will be provided.

The following analyses will be conducted for ICE:

- Overall summary of ICE by worst CTC grade (any grade, grade 3-4, grade 5) presented by PT
- Overall summary of any serious ICE by worst CTC grade (any grade, grade 3-4, grade 5) presented by PT
- Overall summary of any ICE leading to discontinuation of study therapy by worst CTC grade (any grade, grade 3-4, grade 5) presented by PT
- Overall summary of any ICE leading to discontinuation of NKTR-214 by worst CTC grade (any grade, grade 3-4, grade 5) presented by PT. This summary will be conducted for Arm A patients only.
- Time-to onset of ICE (any grade, grade 3-5)

Time-to onset analyses are restricted to treated patients who experienced at least one ICEs.

A by-patient listing of ICE will be provided.

8.3.2 Other Safety Observations (OSO) for NKTR-214

Other safety observations (OSO) for NKTR-214 consists of the following:

- IL-2 mediated AEs
    - Flu-like symptoms
    - Rash and pruritus
    - Fatigue/asthenia
    - Arthralgia
    - Elevated hepatic transaminases
    - Elevated serum creatinine
    - Eosinophilic disorders
    - Hypotension
    - Syncope
- Infusion-related reaction (IRR, same day or next day post study drug administration)
- Cytokine release syndrome/cytokine storm (CRS)
- Capillary leak syndrome (CLS)
- Tachyarrhythmias (including supraventricular and ventricular)

All analyses will be conducted in the Safety I/P-risk population and the Safety all-risk population by treatment arm using the treatment-emergent period.

Table 6 lists the planned analyses for each category of OSO for NKTR-214.

1) Overall summary of OSO by worst CTC grade presented by category/PT (any grade, grade 3-4, grade 5).
2) Overall summary of serious OSO by worst CTC grade presented by category/PT (any grade, grade 3-4, grade 5).
3) Overall summary of OSO leading to discontinuation of any study drug by worst CTC grade presented by category/PT (any grade, grade 3-4).
4) Overall summary of OSO leading to discontinuation of NKTR-214 by worst grade presented by category/PT (any grade, grade 3-4). This summary will be conducted for Arm A patients.
5) Time-to onset of OSO (any grade, grade 3-5) by category
6) Overall summary of OSO by Cycle by Worst Grade by category (any grade, grade 3-5) (for Arm A only, and cycle is every 3 weeks in this analysis)

Time-to onset analyses are restricted to treated patients who experienced at least one OSO in the specific category.

For IRR, events that start after study drug administration on the same day or next day will be included. If AE onset time is available, the time information will be used to exclude events that start before study drug administration; otherwise, AEs that start on the same day as study drug administration will be included. Analysis on IRR will only be performed for Arm A.

A by-patient listing of OSO will be provided.

Table 6: Planned Analyses for OSO for NKTR-214

Table Numbers[a]	Tachyarrhythmias	Syncope	CRS CLS	Infusion-related reaction (same day or next day post study drug administration) (For Arm A only)	Flu-like symptoms; Rash and pruritus; Fatigue/asthenia; Arthralgia; Elevated hepatic transaminases; Elevated serum creatinine; Eosinophilic disorders; Hypotension
1), 2)	Yes	No	No	Yes	Yes
3), 4)	Yes	No	No	Yes	Yes
5)	Yes	Yes	Yes	Yes	Yes
6)	No	Yes	No	Yes	Yes

Abbreviations: CLS = capillary leak syndrome; CRS = cytokine release syndrome/cytokine storm

a. Table descriptions provided in Section 8.3.2.

8.4 Deaths

Deaths will be summarized by treatment group:

- All deaths, reasons for death
- Deaths within treatment-emergent period, and reasons for death
- Deaths within extended treatment-emergent period, and reasons for death

A by-patient listing of deaths will be provided for all randomized patients (ITT I/P-risk population and ITT all-risk population).

8.5 Laboratory Parameters

The analysis population for each laboratory test is restricted to patients in the Safety I/P-risk population and the Safety all-risk population who had both a baseline and at least one treatment-emergent measurement of that laboratory test.

8.5.1 Hematology

The following will be summarized by treatment group as baseline CTC grade and worst CTC grade on-treatment per patient: hemoglobin, platelet, white blood counts (WBC), absolute neutrophils count (ANC) and lymphocyte count (LYMP). A table presenting number and percentage of patients with post-baseline CTC grade increased from baseline (any grade, grade 3-4) by treatment group will be provided.

The analyses will be conducted using the treatment-emergent period.

A by-patient listing of these laboratory parameters will be provided.

Eosinophil (EOS)

- Summary of baseline and highest post-baseline EOS will be provided by treatment group:
  - Mean (standard deviation [SD] and standard error [SE]), median, min and max
  - Percent exceeding thresholds: $\geq 10 \times 10^9$/L, $\geq 5 \times 10^9$/L, $\geq 1.5 \times 10^9$/L
  - Time to highest post-baseline EOS: mean (standard deviation [SD]), median, min, max
  - Duration (Weeks) of Sustained Eosinophils $\geq 1.5 \times 10^9$/L: % of patients with eosinophils returned, % patients with duration longer than 6 months, median and 95% CI will be estimated using the KM method

Plot of mean eosinophil (± SE) by treatment arm and time.

8.5.2 Serum Chemistry

The following will be summarized by treatment group as baseline CTC grade and worst CTC grade on-treatment per patient: ALT, AST, alkaline phosphatase (ALP), total bilirubin, creatinine, amylase and lipase. A table presenting number and percentage of patients with post-baseline abnormalities increased from baseline (any grade, grade 3-4) by treatment group will be provided.

The analyses will be conducted using the treatment-emergent period.

A by-patient listing of these laboratory parameters will be provided.

8.5.3 Electrolytes

The following will be summarized by treatment group as baseline CTC grade and worst CTC grade on-treatment per patient: sodium (high and low), potassium (high and low), calcium (high and low), and glucose. A table presenting number and percentage of patients with post-baseline abnormalities increased from baseline (any grade, grade 3-4) by treatment group will be provided.

The analyses will be conducted using the treatment-emergent period.

A by-patient listing of these laboratory parameters will be provided.

8.5.4 Additional Analyses

In addition, further analyses on specific laboratory parameters will be performed by treatment group:

Abnormal Hepatic Function Test

The number of patients with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- ALT or AST > 3 x ULN, > 5 x ULN, > 10 x ULN and > 20 x ULN
- Total bilirubin > 2 x ULN
- alkaline phosphatase (ALP) > 1.5 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 1.5 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN
- Concurrent (within 30 days) ALT or AST > 3 x ULN and total bilirubin > 2 x ULN and all ALP measurements normal or < 2 x ULN

For identification of concurrent elevations, window for identification of total bilirubin elevation applies to on or after ALT/AST elevation. Window for ALP applies to both before and after ALT/AST elevation.

The analyses will be conducted using the extended treatment-emergent period.

A by-patient listing of these specific abnormalities will be provided.

Abnormal Thyroid Function Test

The number of patients with the following laboratory abnormalities from on-treatment evaluations will be summarized by treatment group:

- TSH value > ULN and
  - with baseline TSH value ≤ ULN
  - with at least one FT3/FT4 test value < LLN within 2-week window after the abnormal TSH test
  - with all FT3/FT4 test values ≥ LLN within 2-week window after the abnormal TSH test
  - with FT3/FT4 missing within 2-week window after the abnormal TSH test.
- TSH < LLN and
  - with baseline TSH value ≥ LLN
  - with at least one FT3/FT4 test value > ULN within 2-week window after the abnormal TSH test
  - with all FT3/FT4 test values ≤ ULN within 2-week window after the abnormal TSH test
  - with FT3/FT4 missing within 2-week window after the abnormal TSH test

The analyses will be conducted using the treatment-emergent period.

A by-patient listing of these specific abnormalities will be provided.

8.6 Additional Safety Analyses

8.6.1 Vital Signs

Values at each scheduled visit, change, and percent change from baseline for vital signs, including systolic blood pressure, diastolic blood pressure, pulse, respiration rate, and body temperature will be summarized descriptively.

Criteria for clinically notable vital sign abnormalities are defined in Table 7 below. Abnormal values for patients exhibiting clinically notable vital sign abnormalities will be listed along with their assigned treatment group. Alert vital signs are defined as an absolute value outside the defined range and percentage change > 25%. The abnormal values for patients exhibiting alert vital sign abnormalities will be listed.

Table 7: Criteria for Clinically Notable Vital Sign Abnormalities

Parameter	High Threshold	Low Threshold
Systolic blood pressure	> 160 mmHg	< 90 mmHg
Diastolic blood pressure	> 100 mmHg	< 60 mmHg
Pulse	> 100 bpm	< 60 bpm
Temperature	> 38ºC	< 35ºC
Respiratory rate	> 24/min	< 12/min

9.0 COVID-19 RELATED ANALYSES

9.1 Study Conduct

9.1.1 Patients Impacted

A brief high-level summary of patients impacted by coronavirus disease 2019 (COVID-19) will be provided by treatment group for the Safety IMDC I/P-risk Analysis Set and Safety IMDC All-risk Analysis Set. This summary will present the number and percentage of patients with any incidence in each of the categories listed below due to COVID-19 and any reason, where total number of patients in corresponding treatment group will be used as denominator for percentage calculation:

- Any positive test of COVID-19
- Received COVID-19 vaccine (type/manufacturer of vaccine, e.g., RNA vaccine from Pfizer) during treatment
- Missed study visit
- Missed at least 9 weeks of doses (For Arm A, missing 2 consecutive doses for NKTR-214 and nivolumab means no infusion in 9 weeks, which is comparable to missing 9 weeks of doses for Arm B).Delayed scan
- Missed scan (any missed scan, 2 scans missed, ≥ 3 scans missed)
- Study treatment discontinuation
- Study discontinuation
- Death
- Study treatment discontinuation

9.1.2 Multiple Incidences

Number of unique incidences per person in each of the categories listed below due to COVID-19 will be summarized by country and by treatment group using descriptive statistics:

- Missed study visit
- Missed at least 9 weeks of doses Delayed scan
- Missed scan

A table showing total number of occurrences (unique incidences) in each category by treatment group will be provided.

A listing of all COVID-19 impact data collected will be created.

9.2 Impacts on Efficacy Results

Supplementary analyses for ORR and OS may be performed to estimate the treatment effects without the COVID-19 impact if needed.

9.3 Impacts on Safety Results

By-patient listing for COVID-19 related adverse events will be provided for the Safety Analysis Set. If sample size allows (e.g., > 10 per treatment group), COVID-19 related TEAEs can be summarized by worst CTC grade and treatment group for COVID-19 infected patients. Impact of COVID-19 vaccines on study drug dosing and adverse events may be further explored.

If central clinical laboratory data was missing for some patients in some visits due to COVID-19, local laboratory data may be collected in place of missed central laboratory data. In such a case, central laboratory and local laboratory data may be combined for the lab related analysis. Specifical considerations are described as follows:

1. For summary of post-baseline laboratory toxicity worsening from baseline

   All (scheduled and unscheduled) central and local laboratory data will be included. The central and local laboratory data will be graded according to their corresponding laboratory normal ranges (central laboratory normal range for central laboratory data, local laboratory normal range for local laboratory data). If the baseline value is involved in the derivation of toxicity grade based on CTCAE 5.0, the corresponding baseline value for central and local laboratory will be used as reference (central laboratory toxicity grade based on central laboratory baseline value, local laboratory toxicity grade based on local laboratory baseline value). If a corresponding baseline value is not available, for that particular laboratory value, the toxicity grade will be derived assuming the baseline value is normal, thus the grade will be derived based on corresponding normal range only.

   In summary table, the baseline toxicity grade will be based on central laboratory data if the central laboratory data is available; otherwise (central laboratory data is not available), then the baseline grade will be based on local laboratory if available. The post-baseline worst toxicity grade will be based on the worst grade of both central and local laboratory data.

2. For summary of abnormal hepatic fuction and summary of abnormal thyroid function

   All (scheduled and unscheduled) central and local laboratory data will be included. The central and local laboratory data (AST, ALT, total bilirubin, TSH, FT3, FT4) will be summarized according to their corresponding laboratory normal ranges (central laboratory normal range for central laboratory data, local laboratory normal range for local laboratory data).

10.0 EFFICACY INTERIM AND FINAL ANALYSES

This study will have 1 planned efficacy interim analysis and 1 final analysis:

- Interim Analysis (IA – ORR & OS interim and OS events re-estimation): when at least 156 OS events in IMDC Intermediate- or Poor-risk Population have been observed.
- Final Analysis (FA – OS final): timing will be determined at the time of IA using the promising zone adaptive method (Mehta and Pocock 2011) to estimate the OS events needed.

11.0 DATA MONITORING COMMITTEE

An external multidisciplinary Data Monitoring Committee (DMC) will be convened for this study. The DMC will be comprised of qualified clinicians and a biostatistician, all independent from the Sponsor and (for the clinicians) independent of investigational sites, selected to avoid conflict(s) of interest. The DMC's specific activities will be detailed in a mutually agreed upon charter, which will define the relevant processes, including meeting proceedings and structure, data assessments, documentation and recordkeeping, process for DMC recommendations, and regulatory reporting, as applicable. The charter will contain procedures to ensure the minimization of bias, such as maintaining confidentiality of any interim data.

The DMC will review the results of the interim analyses as well as assess accumulating safety data and emerging risk/benefit balance at regular intervals and on an ad-hoc basis as detailed in the DMC charter. The DMC will also conduct the OS events re-estimation; details will be provided in the DMC charter.

12.0 PHARMACOKINETIC, PHARMACODYNAMIC, [redacted] ANALYSIS

In addition to the objectives described in this section, PK data obtained in this study may also be combined with data from other studies to develop population PK models. These models may be used to evaluate the effects of intrinsic and extrinsic covariates on the PK of NKTR-214 related molecules and/or nivolumab and to determine measures of individual exposure for nivolumab (such as single dose and steady state peak, trough and time averaged concentration, CL volume, half-life) and for NKTR-214 (such as AUC, clearance, volume, half-life, C_{max}, C_{min}, and C_{avg}). Model predicted exposures may be used for exposure-response analyses of selected efficacy and safety endpoints. Results of any population PK or exposure-response analyses will be reported separately.

[redacted]

12.1 Pharmacokinetic Analysis

Listings for blood collection dates, times and concentrations will be generated by patient for each analyte. Sampling time deviations will be computed as differences between scheduled (nominal) and actual sampling times and expressed in hours and as a percentage of the nominal time.

Observed plasma NKTR-214-RC, NKTR-214-AC, and Total-PEG, and serum nivolumab concentrations at each nominal PK sampling time will be listed by patient and tabulated for each analyte using descriptive statistics, including but not limited to n (observed), n (non-zero), arithmetic mean, standard deviation (SD), standard error of the mean (SE), arithmetic coefficient of variation (CV%), geometric mean, geometric CV%, median, minimum, and maximum.

Mean (+ standard deviation) observed plasma NKTR-214-RC, NKTR-214-AC, and Total-PEG concentrations versus nominal time plots will be generated on linear and semi-logarithmic scales for each analyte for Cycle 1 based on nominal sampling times. For serum nivolumab, mean (+ standard deviation) serum concentrations versus nominal time will be generated on linear scale by nominal sampling time. Nivolumab concentrations should not be connected by a line as all the samples are pre-dose concentrations from different cycles.

PK samples that were collected outside the protocol accepted time windows will be excluded from summary tables and summary figures.

Rounding for the reporting of descriptive statistics will be to three significant digits, except for n, which will be rounded to the whole value. Percentages such as CV% presented in tables will be rounded to one decimal place.

Descriptive statistics will not be performed if $n < 3$ and summary tables will only include median, minimum, and maximum.

12.1.1 Handling of Missing Values, BLQ, and Outliers

- NKTR-214-RC, NKTR-214-AC, and Total-PEG

When summarizing concentration values, below the limit of quatification (BLQ) values will be considered as zero, and zero values will be excluded from the calculation of geometric means and CV% geometric mean, however, zero values will be included for all other summary statistics and the number of non-zero concentrations will be reported. Listings for concentrations values (except for pre-dose C1D1) will list these concentrations as < lower limit of quantification (LLOQ).

Samples with concentrations 5 x higher than mean + standard deviations (SD) for all samples at that time will be considered outliers and not included in analyses and summary statistics and plots. The reason for not using selected data will be presented and discussed in the clinical study report (CSR).

- Nivolumab

When summarizing concentration values, BLQ values (except for pre-dose C1D1) will be imputed as LLOQ/2. Listings for concentrations values (except for pre-dose C1D1) will list these concentrations as <LLOQ. Pre-dose C1D1 BLQ values will be set to zero, and zero values will be excluded from the calculation of geometric means and CV% geometric mean, however, zero values will be included for all other summary statistics and the number of non-zero concentrations will be reported for pre-dose C1D1.

12.2 [redacted]/Pharmacodynamic Analysis

Observed plasma sCD25 and cytokine concentrations, and fold change from baseline at each nominal PK sampling time and maximum fold change and time of maximum fold change will be listed by patient and tabulated for each analyte using descriptive statistics, including but not limited to n (number of non-missing data), arithmetic mean, standard deviation, SE, arithmetic CV%, geometric mean, geometric CV%, median, 25^{th} and 75^{th} percentiles, minimum, and maximum.

For sCD25, all Cycle 1 pre-dose concentrations that are BLQ will be set to 0, all other concentrations that are BLQ will considered missing. Zero values will be excluded from the calculation of geometric means and CV% geometric mean.Missing concentration values will not be included in the descriptive statistics.

For each cytokine, all BLQ values will be set to the lower limit of quantification (LLOQ).

Plot of mean (+SE) plasma sCD25 and cytokine concentrations versus nominal time will be generated for each analyte for baseline and Cycle 1. Fold change from baseline plots will also be generated. Baseline is the last observation prior first dose of study drug within 2 weeks (or 14 days).

Plot of mean (+SE) cytokine concentrations and its fold change from baseline will be generated for baseline and all cycles.

Absolute eosinophil count, absolute lymphocyte count, and their ratios of C1D8 / baseline will be summarized with descriptive statistics: n (number of non-missing data), arithmetic mean, standard deviation, SE, CV%, median, 25^{th} and 75^{th} percentiles, minimum, and maximum.

Plot of mean (+SE) absolute lymphocyte counts and absolute eosinophil counts versus nominal time will be generated for baseline and Cycle 1. Fold change from baseline plots will also be generated. Baseline is the last observation prior to first dose within 2 weeks (or 14 days).

Plot of mean (+SE) eosinohpil trough count and its change from baseline will be generated for baseline and all cycles.

Rounding for the reporting of parameters will be to 3 significant digits. Percentages presented in tables such as CV% will be rounded to one decimal place. The same convention will be followed for descriptive statistics, except for n, which will be rounded to the whole value. Descriptive statistics will not be performed if $n < 3$ and summary tables will only include median, minimum, and maximum.

13.0 DOCUMENT HISTORY

Version Number	Author(s)	Description
1.0	[redacted]	Initial version
2.0	[redacted]	• Section 6.2: Added PFS to testing procedures in Figure 1 • Section 7.2: removed "Only when the overdose is affiliated with adverse events, will this be considered an important protocol deviaton" from the important protocol deviation criterion "Relative dose intensity >125% for study drug (programmatically derived). " • Section 7.1: added "Baseline liver mets as per BICR (Yes, No)" and "number of disease sites per patient for target lesions'. • Section 7.4.3: removed "The number and percent of patients who took at least one subsequent post-study immune-oncology therapy, and post-study TKI therapy will be calculated and presented." • Section 7.5.2: added sensitivity analysis of PFS accounting for two or more missing tumor assessments prior to PFS event (death or documented progression). • Section 7.5.4.3: Removed the sensitivity analysis of restricted mean of duration of response (this analysis will be done for DMC only). • Section 7.8.1: Removed sample level ADA analysis by analyte and by visit. Clarified that samples confirmed to be positive for anti-NKTR-214/anti-PEG, anti-IL-2 antibodies will be tested for neutralizing activity for IL-2 using a validated cell-based assay. Anti-nivolumab antibodies positive samples will also be tested for neutralizing activity for nivolumab using a validated cell-based assay. • Section 7.8.2: Clarified that infusion related reaction events that start after study drug administration on the same day or next day will be included. Clarified that pre-dose NKTR-214 RC and pre-dose NKTR-214 AC concentrations will be summarized by ADA status.

		• Section 8.1: add the following NKTR-214-related TEAE summaries:
		  o NKTR-214-related TEAEs presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT
		  o Grade 3 or above NKTR-214-related TEAE by PT
		  o Serious NKTR-214-related TEAE by PT
		  o NKTR-214-related TEAE leading to death by PT
		  o NKTR-214-related TEAE presented by worst grade (any grade, grade 3-4, grade 5) by SOC and PT (using an extended treatment-emergent period)
		• Section 8.2: Clarified that IMAEs will be summarized using the extended treatment-emergent period for each immune-mediated category for Arm A only. Re-challenge analysis is removed.
		• Section 8.3: Time to resolution analysis is removed.
		• Section 8.3.1: exposure adjusted analysis for ICE is removed.
		• Section 8.6.1: Table 7 (Criteria for Clinically Notable Vital Sign Abnormalities) is updated.
		• Section 9.2: The analysis on COVID 19 impacts on efficacy results may be performed if needed, and sections 9.2.1 (supplemental analysis on objective response rate) and 9.2.2 (supplemental analysis on overall survival) are removed.
		• Section 13: Document History section is added.

14.0 REFERENCES

Clopper, C.; Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934; 26: 404–413.

Greenwood, M. The errors of sampling of the survivorship tables, Reports on Public Health and Statistical Participants, 33, Appendix 1, HMSO, London, 1926

Harrington B, Fleming T. A Class of Rank Test Procedures for Censored Survival Data. Biometrika, Volume 69, Issue 3, 1982.

Huang B, Tian L, et al. Analysis of Response Data for Assessing Treatment Effects in Comparative Clinical Studies. Annals of Internal Medicine 2020;173:368-374.

W. Lehmacher, G. Wassmer, Adaptive sample size calculations in group sequential trials, Biometrics. 1999, 55:1286–1290.

Liu Y, and Hu M. Testing multiple primary endpoints in clinical trials with sample size adaptation. Pharmaceut. Statist. 2016, 15: 37–45. doi: 10.1002/pst.1724.

Mehta C, Pocock SJ. Adaptive increase in sample size when interim results are promising: A practical guide with examples. Statistics in Medicine 2011; 30: 3267-3284.

Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. N Engl J Med. 2018b; 378(14):1277-90.

Rothrock E. Development and Initial Validation of the NCCN/FACT Symptom Index for Advanced Kidney Cancer. Value Health 2013 Jul-Aug; 16(5): 789-796.

The EuroQol Group: EuroQol: A new facility for the measurement of health-related quality

of life - The EuroQol Group. Health Policy 16:199-208, 1990.

Wassmer G. Planning and analyzing adaptive group sequential survival trials. Biometrical Journal. 2006; 48: 714-729.

15.0 APPENDIX

15.1 The Logrank Test for OS Analysis

Let T_1 denote the interim logrank test statistic with d_1 number of events and T_2 denote the logrank test statistic based on the cumulative data at final analysis with adapted d_2^* number of events. Then we calculate the incremental test statistic as following:

$\tilde{T}_2^* = \frac{\sqrt{d_2^*}T_2 - \sqrt{d_1}T_1}{\sqrt{d_2^* - d_1}}$.

The final combination test statistic (Lehmacher and Wassmer, 1999) with weight $\sqrt{\frac{2}{3}}$ and $\sqrt{\frac{1}{3}}$ can be written as $T_2^* = \sqrt{\frac{2}{3}}T_1 + \sqrt{\frac{1}{3}}\tilde{T}_2^*$. The overall Type I error rate is the probability of rejecting OS at either interim or final analysis under the null that there is no difference in survival between the treatment and control group (denoted as H_0). The critical value for the interim analysis is denoted as $z_{\frac{0.01}{2}} = \Phi^{-1}(1 - \frac{0.01}{2})$ (where Φ denotes the CDF for standard normal distribution) as two-sided 0.01 level of alpha is spent at this time. The critical value c for the final analysis is calculated so that the overall type I error rate is maintained at 0.049 (two-sided) or 0.0245 (one-sided). The probability can be calculated as follows:

$$0.0245 = P_{H_0}(T_1 > z_{0.01/2} \text{ or } T_2^* > c) = P_{H_0}(T_1 > z_{0.01/2}) + P_{H_0}(T_1 \leq z_{0.01/2} \text{ and } T_2^* > c)$$

Therefore, we need to find c such that $P_{H_0}\left(T_1 \leq z_{\frac{0.01}{2}} \text{ and } T_2^* > c\right) = 0.0195$. Because T_1 and T_2^* asymptotically follows standard normal distribution with correlation $\sqrt{\frac{2}{3}}$ (Wassmer, 2006), the critical value c can be found to be $z_{0.0231}$. This implies that the significant p-value will be 0.0231 (one-sided) or approximately 0.0462 (two-sided).

With similar arguments, if the overall type I error rate is to be maintained at 0.05 (two-sided) or 0.025 (one-sided), the nominal alpha for OS final analysis with weighted combination test will be 0.0236 (one-sided) or approximately 0.0472 (two-sided).

15.2 The Weighted Logrank Test for PFS Analysis

Let Z denote the weighted logrank test statistic with d_i number of events and Y_i number at risk. The test statistic Z can be written as

$$Z = \frac{\sum_{i=1}^{D} W(t_i)\left[d_{i1} - Y_{i1}\left(\frac{d_i}{Y_i}\right)\right]}{\sqrt{\widehat{Var}(Z)}}$$

where $\widehat{Var}(Z) = \sum_{i=1}^{D} W(t_i)^2 \frac{Y_{i1}}{Y_i}\left(1 - \frac{Y_{i1}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i$ and $W(t_i) = \{\hat{S}(t_{i-1})\}^{\rho}\{1 - \hat{S}(t_{i-1})\}^{\gamma}$.

Considering the expected delayed effect on PFS, we select $\rho = 0$ and $\gamma = 1$ for the Fleming-Harrington test. Therefore, the weight $W(t_i) = 1 - \hat{S}(t_{i-1})$.

The PFS will be tested hierarchically given statistical significance of OS in both IMDC intermediate- or poor-risk (I/P-risk) population and all-risk population. The PFS interim and final analysis will be performed at the same time of OS interim or final analysis. The combination test statistic with weight $\sqrt{\frac{2}{3}}$ and $\sqrt{\frac{1}{3}}$ will be used in the final PFS analysis. For PFS, a 0.03 (two-sided) level of alpha will be spent at interim analysis. Following the same arguments as in Section 15.1, if the overall type I error is 0.049 (two-sided), the nominal alpha for the PFS final analysis will be 0.0323 (two-sided); if the overall type I error is 0.05 (two-sided), the nominal alpha for the PFS final analysis will be 0.0335 (two-sided).

15.3 Proof of Strong Family-wise Error Rate Control for the Four Primary Hypotheses and Two Key Secondary Hypotheses

15.3.1 Study Design

Figure 2 gives study design flow charts for ORR, OS, and Figure 1 shows the overall testing procedure for ORR, OS and PFS where PFS will be tested in IMDC I/P-risk and all-risk populations sequentially after statistical significance of OS in both populations. All p-values and testing levels are two-sided in this section and one-sided in Section 15.3.3. At interim analysis, if OS in IMDC I/P-risk population is not statistically significant at 0.01 level, then number of events in this population for the final analysis will be re-estimated based on the pre-specified adaptation rule. The final analysis will be triggered by the required number of events in the IMDC I/P risk population. At final analysis, if OS in IMDC I/P risk population is statistically significant (at 0.0462 level if ORR is not significant at IA and at 0.0472 level if ORR is significant at IA), then OS in IMDC all-risk population will also be tested at the same level as the IMDC I/P risk population. Given the ststistical significance of OS in the IMDC all-risk population at FA, the key secondary endpoint of PFS by BICR in the IMDC I/P-risk population and the IMDC all-risk population will be test sequentially at level 0.0323 if ORR is not significant at IA and at level 0.0335 if ORR is significant at IA.

On the other hand, if OS in IMDC I/P risk population is statistically significant at 0.01 level at interim, then formal efficacy claim can be obtained, and no further testing will be performed for this population. At the same time, OS in the IMDC all-risk population can also tested at 0.01 level. If statistically significant, the key secondary endpoint of PFS by BICR in both populations will be tested sequentially at 0.03 level. Otherwise, if OS in all-risk population is not significant, number of events in this population for the final analysis will be re-estimated based on the pre-specified adaptation rule. The final analysis occurs when the adapted number of events in the IMDC all-risk population is achieved. The OS in IMDC all-risk population will be tested at 0.0462 level if ORR is not significant at IA and at 0.0472 level if ORR is significant at IA. And the key secondary endpoint of PFS by BICR in the two populations will be tested sequentially at level 0.0323 if ORR is not significant at IA and at level 0.0335 if ORR is significant at IA, given the statistical significance of OS in all-risk population at final analysis.

15.3.2 Notation and Test Statistics

For OS in the IMDC I/P-risk populatio, let T_1 denote the interim logrank test statistic with d_1 number of events and T_2 denote the regular logrank test statistic based on the cumulative data at final analysis with adapted d_2^* number of events. Then we calculate the incremental test statistic following Wassmer (2006):

$$\tilde{T}_2^* = \frac{\sqrt{d_2^*}T_2 - \sqrt{d_1}T_1}{\sqrt{d_2^* - d_1}}.$$

The final weighted test statistic with weight $\sqrt{2/3}$ and $\sqrt{1/3}$ can be written as

$$T_2^* = \sqrt{2/3}\,T_1 + \sqrt{1/3}\,\tilde{T}_2^*.$$

Similarly for OS in the IMDC all-risk population, let T_1^a denote the interim logrank test statistic with d_1^a number of events, let T_2^{a*} denote the logrank test statistic based on the cumulative data at final analysis with d_2^{a*} number of events resulting from adaption in the IMDC I/P population and T_2^{a**} denote the logrank test statistic based on the cumulative data at final analysis with d_2^{a**} number of events resulting from adaption in the IMDC all-risk population. Then the corresponding incremental test statistics could be calculated as follows:

$$\tilde{T}_2^{a*} = \frac{\sqrt{d_2^{a*}}T_2^{a*} - \sqrt{d_1^a}T_1^a}{\sqrt{d_2^{a*} - d_1^a}} \text{ or } \tilde{T}_2^{a**} = \frac{\sqrt{d_2^{a**}}T_2^{a**} - \sqrt{d_1^a}T_1^a}{\sqrt{d_2^{a**} - d_1^a}}.$$

The final test statistics can be written as

$$T_2^{a*} = \sqrt{2/3}T_1^a + \sqrt{1/3}\tilde{T}_2^{a*} \text{ or } T_2^{a**} = \sqrt{2/3}T_1^a + \sqrt{1/3}\tilde{T}_2^{a**}.$$

Following a similar logic, we denote the interim test statistics for PFS in I/P and all-risk populations as Z_1 and Z_1^a, denote the final test statistics for PFS in both populations as Z_2^*, Z_2^{a*} based on adaption in the IMDC I/P population with event size d_2^{a*}, and Z_2^{**} and Z_2^{a**} based on adaption in the IMDC all-risk population with event size d_2^{a**}.

Further, let $H_{01}^{OS}$and H_{02}^{OS} denote the null hypotheses regarding the OS in IMDC I/P-risk and IMDC all-risk population, H_{01}^{ORR} and H_{02}^{ORR} denote the null hypotheses regarding the ORR in IMDC I/P-risk and IMDC all-risk population, and H_{01}^{PFS} and H_{02}^{PFS} denote the null hypotheses regarding the PFS in IMDC I/P-risk and IMDC all-risk population.

15.3.3 Strong Control of Family-wise Error Rate

By the partition principle, the family-wise error rate is strongly controlled at one-sided level 0.025 if the probability of false rejection in each of the following partitioned parameter space could be controlled at the same one-sided 0.025 level. Although there are six individual hypotheses: $H_{01}^{OS}, H_{02}^{OS}, H_{01}^{ORR}, H_{02}^{ORR}, H_{01}^{PFS}$, and H_{02}^{PFS}, due to the decision path between $H_{01}^{OS}, H_{02}^{OS}, H_{01}^{PFS}, H_{02}^{PFS}$ and the decision path between H_{01}^{ORR} and H_{02}^{ORR}, there are only a total of 3*5-1=14 partitioned parameter spaces (Liu and Hsu 2009).

Partitioned parameter space (where $\cdot^C$ denotes the complement of the null space)	***False rejection***
1. $\boldsymbol{H_{01}^{ORR} \cap H_{01}^{OS}}$	Reject H_{01}^{ORR} or H_{01}^{OS}
2. $\boldsymbol{(H_{01}^{ORR})^C \cap H_{02}^{ORR} \cap H_{01}^{OS}}$	Reject H_{02}^{ORR} or H_{01}^{OS}
3. $\boldsymbol{H_{01}^{ORR} \cap (H_{01}^{OS})^C \cap H_{02}^{OS}}$	Reject H_{01}^{ORR} or H_{02}^{OS}
4. $\boldsymbol{(H_{01}^{ORR})^C \cap H_{02}^{ORR} \cap (H_{01}^{OS})^C \cap H_{02}^{OS}}$	Reject H_{02}^{ORR} or H_{02}^{OS}
5. $\boldsymbol{(H_{01}^{ORR})^C \cap (H_{02}^{ORR})^C \cap H_{01}^{OS}}$	Reject H_{01}^{OS}
6. $\boldsymbol{(H_{01}^{ORR})^C \cap (H_{02}^{ORR})^C \cap (H_{01}^{OS})^C \cap H_{02}^{OS}}$	Reject H_{02}^{OS}
7. $\boldsymbol{H_{01}^{ORR} \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap H_{01}^{PFS}}$	Reject H_{01}^{ORR} or H_{01}^{PFS}
8. $\boldsymbol{(H_{01}^{ORR})^C \cap H_{02}^{ORR} \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap H_{01}^{PFS}}$	Reject H_{02}^{ORR} or H_{01}^{PFS}
9. $\boldsymbol{(H_{01}^{ORR})^C \cap (H_{02}^{ORR})^C \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap H_{01}^{PFS}}$	Reject H_{01}^{PFS}
10. $\boldsymbol{H_{01}^{ORR} \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap (H_{01}^{PFS})^C \cap H_{02}^{PFS}}$	Reject H_{01}^{ORR} or H_{02}^{PFS}
11. $\boldsymbol{(H_{01}^{ORR})^C \cap H_{02}^{ORR} \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap (H_{01}^{PFS})^C \cap H_{02}^{PFS}}$	Reject H_{02}^{ORR} or H_{02}^{PFS}
12. $\boldsymbol{(H_{01}^{ORR})^C \cap (H_{02}^{ORR})^C \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap (H_{01}^{PFS})^C \cap H_{02}^{PFS}}$	Reject H_{02}^{PFS}
13. $\boldsymbol{H_{01}^{ORR} \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap (H_{01}^{PFS})^C \cap (H_{02}^{PFS})^C}$	Reject H_{01}^{ORR}
14. $\boldsymbol{(H_{01}^{ORR})^C \cap H_{02}^{ORR} \cap (H_{01}^{OS})^C \cap (H_{02}^{OS})^C \cap (H_{01}^{PFS})^C \cap (H_{02}^{PFS})^C}$	Reject H_{02}^{ORR}

- Under partitioned parameter spaces 1 or 2 (with j=1 or 2),

  P (false rejection) = P (Reject H_{0j}^{ORR} or H_{01}^{OS})

  = P (Reject H_{0j}^{ORR}) + P (Reject H_{01}^{OS} but do not reject H_{0j}^{ORR})

  $\leq 0.0005 + P\left(T_1 > z_{0.005} \text{ or } T_2^* > z_{0.0462/2}\right)$

  $= 0.0005 + P(T_1 > z_{0.005}) + P\left(T_1 \leq z_{0.005} \text{ and } T_2^* > z_{0.0462/2}\right)$

  =0.0005+0.005+0.0195 = 0.025

  since T_1 and T_2^* asymptotically follows standard bivariate normal distribution with correlation $\sqrt{2/3}$. (Wassmer, 2006)

- Under partitioned parameter space 3 or 4 (with j=1 or 2),

  P (false rejection) = P (Reject H_{0j}^{ORR} or H_{02}^{OS})

  = P (Reject H_{0j}^{ORR}) + P (Reject H_{02}^{OS} but do not reject H_{0j}^{ORR})

  $\leq 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005})$

  $+ P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0462/2})$

  $+ P(T_1 \leq z_{0.005}, T_2^* > z_{0.0462/2}, T_2^{a*} > z_{0.0462/2})$

  $\leq 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005})$

  $+ P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0462/2})$

  $+ P(T_1 \leq z_{0.005}, T_2^{a*} > z_{0.0462/2})$

$\leq 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005})$

$+ P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0462/2})$

$+ P(T_1 \leq z_{0.005}, T_1^a > z_{0.005}) + P(T_1 \leq z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a*} > z_{0.0462/2})$

$= 0.0005 + P(T_1^a > z_{0.005}) + P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0462/2})$

$+ P(T_1 \leq z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a*} > z_{0.0462/2})$

$= 0.0005 + 0.005 + P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, \sqrt{2/3}T_1^a + \sqrt{1/3}\tilde{T}_2^{a**} > z_{0.0462/2})$

$+ P(T_1 \leq z_{0.005}, T_1^a \leq z_{0.005}, \sqrt{2/3}T_1^a + \sqrt{1/3}\tilde{T}_2^{a*} > z_{0.0462/2})$

By Wassmer (2006), the conditional null distribution of $\tilde{T}_2^{a*}$ and $\tilde{T}_2^{a**}$ given the interim data (T_1 and T_1^a) approximately equals standard normal distribution, denoted as $\tilde{T}_2^{a*}|(T_1, T_1^a)$ $\overset{d}{\Leftrightarrow} \tilde{T}_2^{a**}|(T_1, T_1^a) \overset{d}{\Leftrightarrow} T \sim N(0,1)$. Therefore, the following joint distributions are the same: $(T_1, T_1^a, \tilde{T}_2^{a*}) \overset{d}{\Leftrightarrow} (T_1, T_1^a, \tilde{T}_2^{a**}) \overset{d}{\Leftrightarrow} (T_1, T_1^a, \mathrm{T})$. Subsequently, the following joint distributions are also the same:

$(T_1, T_1^a, \sqrt{2/3}T_1^a + \sqrt{1/3}\tilde{T}_2^{a*}) \overset{d}{\Leftrightarrow} (T_1, T_1^a, \sqrt{2/3}T_1^a + \sqrt{1/3}\tilde{T}_2^{a**}) \overset{d}{\Leftrightarrow} (T_1, T_1^a, \sqrt{2/3}T_1^a + \sqrt{1/3}\,\mathrm{T})$(1)

Hence, above= $0.0055 + P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, \sqrt{2/3}T_1^a + \sqrt{1/3}T > z_{0.0462/2})$

$+ P(T_1 \leq z_{0.005}, T_1^a \leq z_{0.005}, \sqrt{2/3}T_1^a + \sqrt{1/3}T > z_{0.0462/2})$.

$= 0.0055 + P(T_1^a \leq z_{0.005}, \sqrt{2/3}T_1^a + \sqrt{1/3}T > z_{0.0462/2})$

$= 0.025$.

- Under partitioned parameter space 5,

  P (false rejection) = P (Reject H_{01}^{OS}) $\leq P(T_1 > z_{0.005}$ or $T_2^* > z_{0.0472/2}) = 0.025$

- Under partitioned parameter space 6,

  P (false rejection) = P (Reject H_{02}^{OS}) $\leq P(T_1 > z_{0.005}, T_1^a > z_{0.005})$

  $+ P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0472/2})$

  $+ P(T_1 \leq z_{0.005}, T_2^* > z_{0.0472/2}, T_2^{a*} > z_{0.0472/2})$

  $\leq P(T_1 > z_{0.005}, T_1^a > z_{0.005}) + P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0472/2})$

  $+ P(T_1 \leq z_{0.005}, T_2^{a*} > z_{0.0472/2})$

  $= P(T_1 > z_{0.005}, T_1^a > z_{0.005}) + P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0472/2})$

  $+ P(T_1 \leq z_{0.005}, T_1^a > z_{0.005}, T_2^{a*} > z_{0.0472/2}) + P(T_1 \leq z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a*} > z_{0.0472/2})$

  $\leq P(T_1 > z_{0.005}, T_1^a > z_{0.005}) + P(T_1 > z_{0.005}, T_1^a \leq z_{0.005}, T_2^{a**} > z_{0.0472/2})$

$+ P(T_1 \le z_{0.005}, T_1^a > z_{0.005}) + P(T_1 \le z_{0.005}, T_1^a \le z_{0.005}, T_2^{a*} > z_{0.0472/2})$

$= P(T_1^a > z_{0.005}) + P(T_1 > z_{0.005}, T_1^a \le z_{0.005}, T_2^{a**} > z_{0.0472/2})$

$+ P(T_1 \le z_{0.005}, T_1^a \le z_{0.005}, T_2^{a*} > z_{0.0472/2})$

(following similar logic in (1)) = $0.005 + P(T_1^a \le z_{0.005}, \sqrt{2/3}T_1^a + \sqrt{1/3}T > z_{0.0472/2})$

=0.025

- Under partitioned parameter spaces 7 or 8 (with j=1 or 2),

P (false rejection) = P (Reject H_{0j}^{ORR} or H_{01}^{PFS})

= P (Reject H_{0j}^{ORR}) + P (Reject H_{01}^{PFS} but do not reject H_{0j}^{ORR})

$\le 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005}, Z_1 > z_{0.015})$

$+ P(T_1 > z_{0.005}, T_1^a \le z_{0.005}, T_2^{a**} > z_{0.0462/2}, Z_2^{**} > z_{0.0323/2})$

$+ P(T_1 \le z_{0.005}, T_2^* > z_{0.0462/2}, T_2^{a*} > z_{0.0462/2}, Z_2^* > z_{0.0323/2})$

$\le 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005}, Z_1 > z_{0.015})$

$+ P(T_1 > z_{0.005}, T_1^a \le z_{0.005}, T_2^{a**} > z_{0.0462/2}, Z_2^{**} > z_{0.0323/2})$

$+ P(T_1 \le z_{0.005}, T_2^{a*} > z_{0.0462/2}, Z_2^* > z_{0.0323/2})$

$= 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005}, Z_1 > z_{0.015})$

$+ P(T_1 > z_{0.005}, T_1^a \le z_{0.005}, T_2^{a**} > z_{0.0462/2}, Z_2^{**} > z_{0.0323/2})$

$+ P(T_1 \le z_{0.005}, T_1^a > z_{0.005}, T_2^{a*} > z_{0.0462/2}, Z_2^* > z_{0.0323/2})$

$+ P(T_1 \le z_{0.005}, T_1^a \le z_{0.005}, , T_2^{a*} > z_{0.0462/2}, Z_2^* > z_{0.0323/2})$

$\le 0.0005 + P(T_1 > z_{0.005}, T_1^a > z_{0.005}, Z_1 > z_{0.015})$

$+ P(T_1 > z_{0.005}, T_1^a \le z_{0.005}, Z_2^{**} > z_{0.0323/2})$

$+ P(T_1 \le z_{0.005}, T_1^a > z_{0.005}, Z_2^* > z_{0.0323/2})$

$+ P(T_1 \le z_{0.005}, T_1^a \le z_{0.005}, Z_2^* > z_{0.0323/2})$

$< 0.0005 + P\big(T_1 > z_{0.005}, T_1^a > z_{0.005}, (Z_1 > z_{0.015}\ or\ Z_2^{**} > z_{0.0323/2})\big)$

$+ P(T_1 > z_{0.005}, T_1^a \le z_{0.005}, (Z_1 > z_{0.015}\ or\ Z_2^{**} > z_{0.0323/2}))$

$+ P(T_1 \le z_{0.005}, T_1^a > z_{0.005}, (Z_1 > z_{0.015}\ or\ Z_2^* > z_{0.0323/2}))$

$+ P(T_1 \le z_{0.005}, T_1^a \le z_{0.005}, (Z_1 > z_{0.015}\ or\ Z_2^* > z_{0.0323/2}))$

$= 0.0005 + P\big(T_1 > z_{0.005}, (Z_1 > z_{0.015}\ or\ Z_2^{**} > z_{0.0323/2})\big)$

$+ P(T_1 \le z_{0.005}, (Z_1 > z_{0.015}\ or\ Z_2^* > z_{0.0323/2}))$

(following similar logic in (1)) = 0.0005+ $P\left(Z_1 > z_{0.015}\ or \sqrt{2/3}Z_1 + \sqrt{1/3}T > z_{0.0323/2}\right)$

=0.025

- Under partitioned parameter spaces 9, 10, 11, or 12 the false rejection rate is controlled by similar arguments as shown for partitioned parameter space 7 or 8.
- Under partitioned parameter spaces 13 or 14 (with j=1 or 2),

  P (False rejection) = P (Reject H_{0j}^{ORR}) = 0.0005 < 0.025

15.4 Time-to onset and time-to resolution definition and conventions for select adverse events, immune-mediated adverse events and other AEs of interest for NKTR-214

15.4.1 Time-to onset definition

Time-to onset of AE (any grade) for a specific category is defined as the time between the day of the first dose of study treatment and the onset date of the earliest AE (of any grade) in this category.

The time-to onset of AE (grade 3-5) for a specific category is defined similarly with an onset date corresponding to a grade 3-5 AE.

Time-to onset of drug-related AE (any grade or grade 3-5) for a specific category is defined similarly but restricted to drug-related AE.

Time-to onset for a specific subcategory is defined similarly but restricted to event of this subcategory.

15.4.2 Time-to resolution definition

In order to derive the time-to resolution, overlapping or contiguous AEs within a specific category or subcategory will be collapsed into what will be termed "clustered" AEs. For example, if a patient (without pre-treatment AE) experienced an AE from 1st to 5th January, another AE (with different PT but within same category) from 6th to 11th January and same AE from 10th to 12th January, these will be collapsed into one clustered AE from 1st to 12th January. Table 8 is summarizing key derivation steps for each type of clustered AEs.

Time-to resolution of AE (any grade) for a specific category is defined as the longest time from onset to complete resolution or improvement to the grade at baseline among all clustered AEs experienced by the patient in this category per adverse event criteria category. Events which worsened into grade 5 events (death) or have a resolution date equal to the date of death are considered unresolved. If a clustered AE is considered as unresolved, the resolution date will be censored to the last known alive date. Improvement to the grade at baseline implies that all different events in the clustered adverse event should at least have improved to the corresponding

(i.e. with same preferred term) baseline grade. This measure is defined only for patients who experienced at least one AE in the specific category.

The time-to resolution of AE (grade 3-5) for a specific category is defined similarly with an onset date corresponding to a grade 3-5 AE.

Time-to resolution of drug-related AE (any grade or grade 3-5) for a specific category is defined similarly but restricted to drug-related AE.

The time-to resolution of AE (any grade or grade 3-5, drug-related or all) where immune modulating medication was initiated is defined similarly. For data presentation not restricted to IMAE, the additional condition that the patient started an immune modulating medication during the longest AE resolution period will be applied.

Time-to resolution for a specific subcategory is defined similarly but restricted to event of this subcategory.

Table 8: Derivation of Clustered AE

Type of clustered AE	**Derivation**
Any grade	Collapse any on-treatment AE from the same category
Drug-related of any grade	Collapse any on-treatment drug-related AE from the same category
Grade 3-5	Collapse any on-treatment AE from the same category. Resolution will be based on the onset date of the earliest grade 3-5 records (if no grade 3-5 record, clustered AE is excluded)
Drug-related of Grade 3-5	Collapse any on-treatment drug-related AE from the same category Resolution will be based on the onset date of the earliest grade 3-5 record (if no Grade 3-5 record, clustered AE is excluded)

The algorithm for collapsing adverse event records is using the following conventions:

For each patient and specified category, the corresponding adverse event records will be collapsed when:

1) Multiple adverse event records have the same onset date.
2) The onset date of an event record is either the same day or 1 day later than the resolution date of a preceding event record (contiguous events).
3) The onset date of an event record is after the onset date and prior to or on the resolution date of a preceding event record (overlapping events).