STATISTICAL ANALYSIS PLAN

FOR CLINICAL STUDY REPORT

A PHASE 3, RANDOMIZED, OPEN-LABEL STUDY OF NIVOLUMAB COMBINED WITH IPILIMUMAB VERSUS SUNITINIB MONOTHERAPY IN SUBJECTS WITH PREVIOUSLY UNTREATED, ADVANCED OR METASTATIC RENAL CELL CARCINOMA

PROTOCOL(S) CA209214

VERSION # 4.0

[redacted]

[redacted]

[redacted]

2.3 Blinding and Unblinding

This is an open label study.

2.4 Protocol Amendments

Not applicable.

2.5 Data Monitoring Committee

An independent Data Monitoring Committee (DMC) has been established to provide oversight of safety and efficacy considerations, study conduct, and risk-benefit ratio. Following review, the DMC will recommend continuation, modification, or discontinuation of this study based on reported safety and efficacy data. Details of DMC responsibilities and procedures are specified in the DMC charter. Representatives of the Sponsor will serve only as coordinators of the committee, without having full member responsibilities or privileges. In addition, the Sponsor will independently review safety data in a blinded manner during the conduct of this trial to ensure that any safety issues are identified and addressed.

The DMC will conduct the first review of the safety data after at least 20 subjects are treated and followed for at least 1 month. The DMC will conduct its second review of the safety data after at least 50 subjects are treated and followed for at least 1 month. The DMC will conduct its third review of the safety data focusing on the initial approximately 12 Japanese subjects treated and followed for at least 1 month. The DMC will then review safety and the available efficacy data pertaining to co-primary endpoints to evaluate safety in the context of benefit, every six months thereafter.

The DMC will also review the formal analysis of ORR (per IRRC) scheduled at around 27 months (when all patients have at least 6 months of follow-up) from FPFV. Details of the formal ORR analyses can be found in section 7.5.13.

The DMC will also review the formal final analysis of PFS (as per IRRC) and first interim analysis of superiority of OS scheduled at around 31 months (approximately 591 PFS events and 370 OS events) from FPFV. A second interim analysis of overall survival will be at around 40 months (approximately 479 OS events) from FPFV. Details of the interim analyses can be found in section 7.5.13.

2.6 Independent Radiological Review Committee

An independent Radiological Review Committee (IRRC) has been established to provide an independent imaging review of images obtained in subjects participating in this study. Details of IRRC responsibilities and processes may be found in the IRRC Charter. The IRRC determined PFS and ORR endpoints will be utilized as a part of primary and secondary efficacy analyses.

3 OBJECTIVES

3.1 Primary

- To describe the ORR of nivolumab combined with ipilimumab and sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC, as assessed by IRRC.
- To compare the PFS of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC, as assessed by IRRC.
- To compare the OS of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC.

3.2 Secondary

- To compare the PFS of nivolumab combined with ipilimumab to sunitinib monotherapy in all randomized subjects with previously untreated mRCC, as assessed by IRRC.
- To compare the OS of nivolumab combined with ipilimumab to sunitinib monotherapy in all randomized subjects with previously untreated mRCC.
- To estimate the objective response rate (ORR) of nivolumab combined with ipilimumab to sunitinib monotherapy in subjects with previously untreated mRCC (any-risk), as assessed by IRRC.

4 ENDPOINTS

The primary objectives of this study are to describe ORR (as assessed by an IRRC) and to compare PFS (as assessed by an IRRC) and OS of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC. This is measured by the three co-primary endpoints defined in section 4.1.

The first secondary objective of this study is to compare PFS (as assessed by an IRRC) in the two treatment arms in the all randomized population. This would be measured by the same definitions of PFS, as specified in sections 4.1.1 and 4.1.3 respectively, in the all randomized population.

The second secondary objective of this study is to compare OS in the two treatment arms in the all randomized population. This would be measured by the same definition of OS, as specified in section 4.1.4, in the all randomized population.

The third secondary objective of this study is assessing ORR in the two treatment arms in all randomized population. This would be measured by the definition of ORR as specified in section 4.2.1, in the all randomized population.

4.1 Co-Primary Endpoints

Objective response rate, overall survival and progression-free survival are the co-primary endpoints.

4.1.1 Objective Response Rate

Objective response rate is defined as the proportion of randomized subjects who achieve a best response of complete response (CR) or partial response (PR) using the RECIST 1.1 criteria based on IRRC assessment. BOR is defined as the best response designation, as determined by the IRRC, recorded between the date of randomization and the date of objectively documented progression per RECIST 1.1 or the date of subsequent therapy (including tumor-directed radiotherapy and

tumor-directed surgery), whichever occurs first. For subjects without documented progression or subsequent therapy, all available response designations will contribute to the BOR assessment. Confirmation of response is required at least 4 weeks after the initial response. Duration of response (DOR) is defined as the time between the date of first documented response (CR or PR) to the date of the first documented progression as determined by the IRRC (per RECIST 1.1), or death due to any cause, whichever occurs first. For subjects who neither progress nor die, the duration of objective response will be censored at the same time they will be censored for the primary definition of PFS. Time to Objective Response (TTR) is defined as the time from randomization to the date of the first confirmed documented response (CR or PR), as assessed by the IRRC. DOR and TTR will be evaluated for responders (confirmed CR or PR) only.

4.1.2 Primary Definition of Progression-free Survival

The primary definition of PFS (PFS truncated at subsequent therapy) is specified as the time between the date of randomization and the first date of documented progression, as determined by the IRRC (as per RECIST 1.1 criteria), or death due to any cause, whichever occurs first.

Subsequent therapy includes anticancer therapy, tumor directed radiotherapy, or tumor directed surgery as shown in Table 4.1.2-1. Subjects 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.

- Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment.
- Subjects who did not have any on study tumor assessments and did not die will be censored on their date of randomization.
- Subjects who receive subsequent systemic anti-cancer therapy prior to documented progression will be censored at the date of the last tumor assessment conducted on or prior to the initiation of the new therapy.
- Subjects who did not have a documented progression and received subsequent anti-cancer therapy will be censored at the date of the last tumor assessment conducted on or prior to the initiation of the new therapy.

The progression free survival rate at time T is defined as the probability that a subject has not progressed and is alive at time T following randomization.

The first on-study tumor assessment is scheduled to be conducted at 12 weeks (± 1 week) following randomization. Subsequent tumor assessments are scheduled every 6 weeks (± 1 week) up to 13 months, then every 12 weeks until disease progression.

Censoring rules for the primary definition of PFS (PFS truncated at subsequent therapy) are presented as follows and in Table 4.1.2-1.

[redacted]

[redacted]

4.1.3 Secondary Definition of Progression-free Survival

The secondary definition of PFS (ITT definition) is defined as the time between the date of randomization and the first date of documented progression, as determined by the IRRC (as per RECIST 1.1 criteria), or death due to any cause, whichever occurs first. Subjects 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 secondary definition of PFS.

- Subjects who did not progress or die will be censored on the date of their last evaluable tumor assessment.
- Subjects who did not have any on study tumor assessments and did not die will be censored on their date of randomization.

The progression free survival rate at time *T* is defined as the probability that a subject has not progressed and is alive at time *T* following randomization.

The first on-study tumor assessment is scheduled to be conducted at 12 weeks (± 1 week) following randomization. Subsequent tumor assessments are scheduled every 6 weeks (± 1 week) up to 13 months, then every 12 weeks until disease progression.

4.1.4 Overall Survival

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

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

4.2 Secondary Endpoint

4.2.1 Objective Response Rate

Objective response rate is defined as the proportion of randomized subjects who achieve a best response of complete response (CR) or partial response (PR) using the RECIST 1.1 criteria as per IRRC assessment.

The ORR (as per IRRC assessment) is defined as the number of subjects whose *confirmed* best objective (BOR) response is a complete response (CR) or partial response (PR) divided by the number of subjects in the population of interest. The BOR is defined as the best response designation, as determined by the IRRC per RECIST 1.1. For subjects without document progression or subsequent therapy, all available response designations will contribute to the BOR determination. Subsequent therapy includes anticancer therapy, tumor directed radiotherapy, or tumor directed surgery. The BOR will be determined based on response designations up to the date of last evaluable tumor assessment prior to initiation of the subsequent therapy. For subjects who continue treatment beyond progression, the BOR will be determined based on response designations up to the time of initial RECIST 1.1 progression.

The first on-study tumor assessment is scheduled to be conducted at 12 weeks (± 1 week) following randomization. Subsequent tumor assessments are scheduled every 6 weeks (± 1 week) up to 13 months, then every 12 weeks until disease progression.

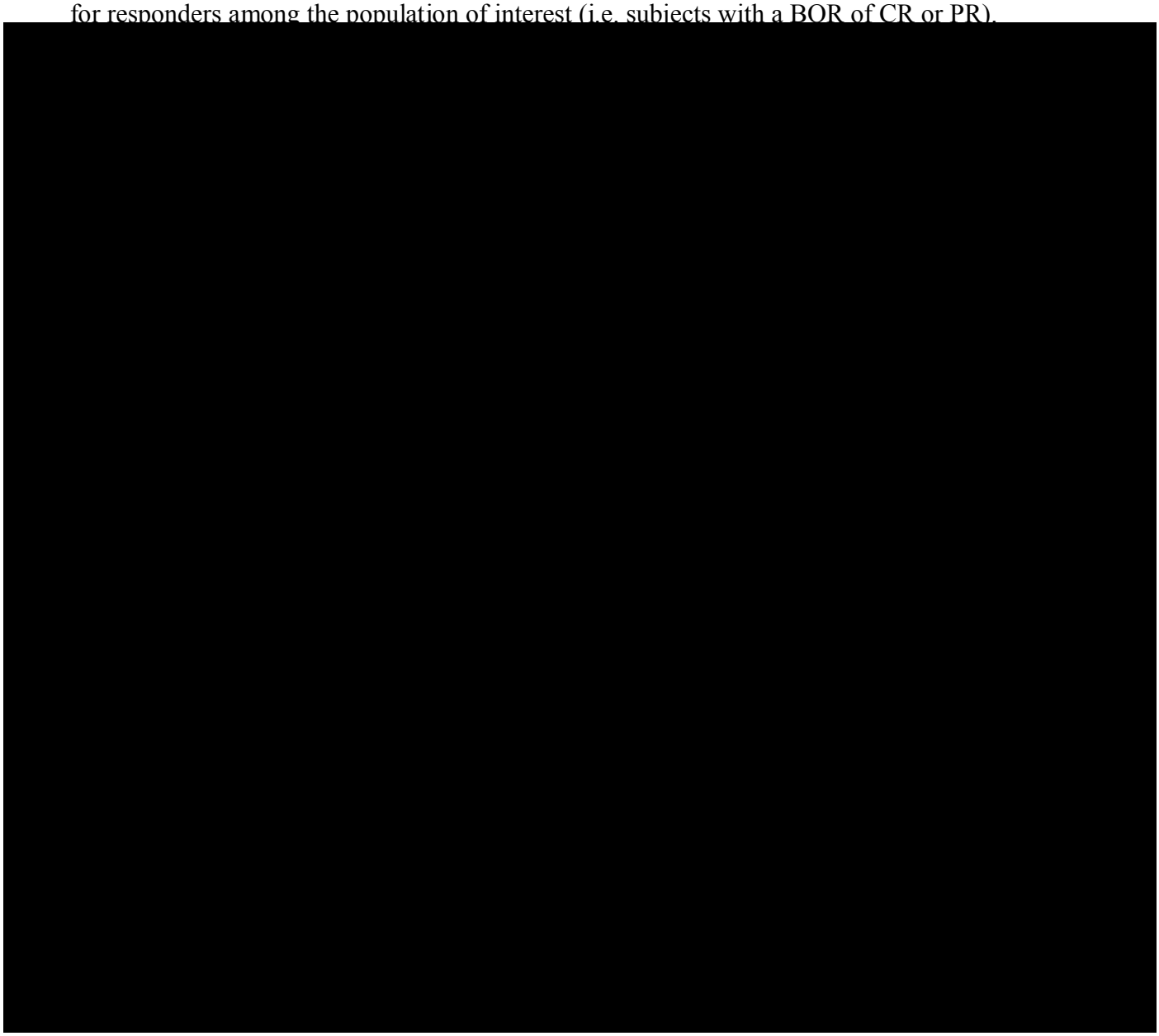
4.2.1.1 Further Characterization of ORR

4.2.1.1.1 Duration of Objective Response

Duration of Objective Response (DOR) is defined as the time between the date of first documented response (CR or PR) to the date of the first documented progression as determined by the IRRC (per RECIST 1.1), or death due to any cause, whichever occurs first. For subjects who neither progress nor die, the duration of objective response will be censored at the same time they will be censored for the primary definition of PFS (Table 4.1.2-1). DOR will be evaluated for responders (i.e. subjects with confirmed CR or PR) only.

4.2.1.1.2 Time to Objective Response

Time to Objective Response (TTR) is defined as the time from randomization to the date of the first confirmed documented response (CR or PR), as assessed by the IRRC. TTR will be evaluated for responders among the population of interest (i.e. subjects with a BOR of CR or PR).



5 SAMPLE SIZE AND POWER

The sample size of the study accounts for the three co-primary efficacy endpoints: ORR and PFS as per IRRC and OS evaluated in intermediate and poor-risk subjects with previously untreated mRCC. The overall alpha for this study is 0.05, which is split with 0.001 to evaluate ORR, 0.009 to evaluate PFS and 0.04 to evaluate OS.

ORR will be analyzed initially on a descriptive basis and will occupy an administrative adjustment of alpha of 0.001. PFS will be evaluated for treatment effect at an alpha of 0.009 (two-sided, penalized 0.001 from a 0.01 allocation), with at least 90% power; no interim analysis of PFS is planned. OS will be evaluated for treatment effect at an alpha level of 0.04 (two-sided) with 90% power, accounting for two formal interim analyses to assess efficacy.

It is estimated that approximately 1070 previously untreated mRCC subjects will be randomized in a 1:1 ratio. Among them, 820 subjects (76.6%) with intermediate/poor risk and approximately 250 (23.4%) subjects with favorable risk as per IMDC (IMDC prognostic score = 0) will be randomized. Assuming a fixed accrual rate of 69 subjects per month (53 intermediate/poor risk

subjects per month), it will take 16 months to randomize 1070 subjects (820 intermediate/poor risk subjects).

Assuming a 21% screen failure rate, it is estimated that approximately 1355 subjects will be enrolled in order to have 820 intermediate/poor-risk subjects randomized. The primary analysis is based on intermediate/poor risk subjects as per IMDC prognostic score and the number of PFS/OS events observed among them. The enrollment will stop once approximately 820 intermediate/poor risk subjects have been randomized.

Sample size justification for ORR estimate

One of the primary objectives of the study is to describe the ORR (based on IRRC assessment) of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC.

The primary analysis of ORR in the intermediate and poor-risk randomized subjects will be performed when these patients have at least 6 months of minimum follow-up from the completion of enrollment. This will allow sufficient follow-up for ORR to have a stable estimate, adequate safety follow-up as well as information on duration of response in this population.

The maximum width of the exact two-sided 95% confidence interval (CI) is 9.9% when the ORR is expected to be in the 20% to 50% range. Table 5-1 summarizes the 95% exact CI when observed ORRs are 20% to 50%, respectively.

Table 5-1: Observed ORR with exact 95% CI

Observed ORR	95% Exact CI
20%	(16.2% - 24.2%)
25%	(21.0% - 29.6%)
30%	(25.6% - 34.7%)
35%	(30.5% - 40.0%)
40%	(35.2% - 44.9%)
45%	(40.2% - 50.1%)
50%	(45.1% - 54.9%)

For example if at least 123 responders are observed among the 410 nivolumab and ipilimumab combination intermediate/poor risk randomized subjects (i.e. ORR ≥ 30%) then the lower bound of the 95% CI is above 25.6%.

Sample size justification for PFS comparison

One of the primary objectives of the study is to compare the progression-free survival (as determined by IRRC) of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC. The number of events and power for this study were calculated assuming an exponential distribution for PFS in each arm.

For this comparison of PFS, it will be required to observe approximately 465 PFS events among the randomized intermediate/poor risk subjects in the two respective treatment arms for a two-sided experiment-wise $\alpha = 0.009$ log-rank test, to show a statistically significant difference in PFS between the treatment arms with approximately 80% power when the true hazard ratio of the experimental arm to control arm is 0.73. The HR of 0.73 is equivalent to demonstrating a 37.8% improvement in median PFS, assuming a median PFS of 9 months in the sunitinib montherapy arm (weighted median estimate assuming a median PFS of 11 months in intermediate risk subjects and median PFS of 4 months in poor risk subjects)[5] and a median PFS of 12.4 months in the experimental treatment arm.

Under the assumptions for accrual and PFS distribution stated above, it will take approximately 35 months from FPFV to observe the required number of PFS events for the final PFS analysis (16 months for accrual and 19 months for minimum follow up). It is projected that an observed HR of 0.785 or less corresponding to a 2.5 month or greater improvement in median PFS (9 vs 11.5 months) for this comparison, would result in a statistically significant improvement in the final analysis of PFS.

Update to Timing of Final PFS Analysis

The number of events for the primary endpoint, PFS per IRRC accounting for subsequent therapy, was observed to be lower than originally assumed per protocol. At 28 months after FPFV there were approximately 72% of the 591 target PFS events per IRRC among intermediate/poor risk subjects, with only 5-10 events occurring monthly over the previous 6 months. This event rate is expected to continue and as a result the target number of events is unlikely to occur even in the next 1-2 years. Thus the timing of the final PFS analysis was advanced with lower power than originally planned as summarized in Table 5-2.

Sample size justification for OS comparison:

One of the primary objectives of the study is to compare the overall survival of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC. The number of events and power of this study were calculated assuming an exponential distribution for OS in each arm.

Approximately 639 events (ie, deaths), observed among the randomized intermediate/poor risk subjects, provides 90% power to detect a hazard ratio (HR) of 0.766 with an overall type 1 error of 0.04 (two-sided). The HR of 0.766 corresponds to a 30.6% increase in the median OS, assuming a median OS of 20 months for sunitinib monotherapy (weighted median estimate assuming a median OS of 26 months in intermediate risk subjects and a median of 8 months in poor risk subjects)[6] and 26.1 months for experimental treatment arms, respectively. It is projected that an observed hazard ratio of 0.846 or less, which corresponds to a 3.6 months or greater improvement in median OS (20 mo vs. 23.6 mo), would result in a statistically significant improvement in OS for the experimental arm at the final OS analysis.

Two formal interim analyses of OS are planned for this study. The first interim analysis is planned at the time of final PFS analysis and it is expected to observe 330 events (52% of the targeted OS

events for final analysis) and the second after observing 479 events (75% of targeted OS events needed for final analysis). The stopping boundaries at the interim and final analyses will be derived based on the number of deaths using O'Brien and Fleming α–spending function.

Under the assumptions stated above on accrual and OS distribution, it will approximately take 65 months from FPFV to observe the required number of OS events for the final OS analysis (16 months for accrual and 49 months for minimum follow up).

In summary, it is expected to take:

- Approximately 16 months to complete accrual
- Approximately 22 months from FPFV to obtain at least a minimum follow-up of 6 months on the intermediate and poor risk randomized subjects for the descriptive analysis of ORR
- Approximately 35 months from FPFV to obtain the approximate number of PFS events (i.e. approximately 465 events among the 820 intermediate and poor risk randomized subjects) and deaths for the first formal interim analysis of OS (i.e. approximately 330 deaths among the same population)
- Approximately 46 months from FPFV to obtain the required deaths for the second formal interim analysis of OS (i.e. 479 deaths among the intermediate and poor risk randomized subjects)
- Approximately 65 months from FPFV to obtain the required deaths for the final analysis of OS (i.e. 639 deaths among the intermediate and poor risk randomized subjects).

Table 5-2 summarizes sample size design parameters and schedule of primary endpoint analyses planned in this study.

Table 5-2: Summary of sample size parameters and schedule of analyse

Co-Primary Endpoints	**ORR**	**PFS**	**OS**
Primary analysis population	Intermediate/poor risk subjects (IMDC score ≥ 1)		
Accrual rate per month	53[b]		
Power	N/A	~80%	90%
Alpha	Administrative 0.001	0.009 2-sided	0.04 2-sided (0.0024 at IA1, 0.0137 at IA2 , 0.0354 at FA)
Hypothesized Median Control vs. exp (months)	25% vs 40%	9 vs. 12.4	20 vs. 26.1
Hypothesized Hazard ratio	N/A	0.726	0.766
Critical Hazard ratio (Observed hazard ratio at which a statistically significant difference would be observed) / Difference in median (months) Corresponding to a minimal clinically significant effect size	N/A	0.785 / 2.5	0.846/ 3.6
Critical HR at interim analysis-1(IA1) /effect size	N/A	N/A	0.72/ 7.8

Table 5-2: Summary of sample size parameters and schedule of analyse			
Co-Primary Endpoints	**ORR**	**PFS**	**OS**
Expected number of event for IA1 (percentage of target events)	N/A	N/A	330 (52%)
Timing of IA1 from FPFV l(months)	N/A	N/A	35
Critical HR at interim analysis-2(IA2) /effect size	N/A	N/A	0.8 / 5.1
Target number of event for IA2 (percentage of target events)	N/A	N/A	479 (75%)
Timing of IA2 from FPFV (months)	N/A	N/A	46
Accrual Duration (months)	16	16	16
Timing of final analysis (FA) from FPFV (months)	22	35	65
Sample size[a]	820	820	820
Target number of events (Event Goal)	N/A	465	639

[a] East version 5.4 was used for sample size / power computation.

[b] Accrual rate adjusted to reflect observed accrual.

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

Baseline period:

- Baseline evaluations or events will be defined as evaluations or events that occur before the date and time of the first dose of study treatment, for all treated subjects. For subjects who are randomized but not treated, baseline evaluation or events will be defined as those that occur before the date and time of randomization.
- In cases where the time (onset time of event or evaluation time and dosing time) is missing or not collected, the following definitions will apply:
  - Pre-treatment AEs will be defined as AEs with an onset date prior to but not including the day of the first dose of study treatment.
  - Baseline evaluations (laboratory tests, pulse oximetry and vital signs) will be defined as evaluations with a date on or prior to the day of first dose of study treatment.

If there are multiple valid assessments, the assessment that is closest to day (and time if collected) of the first dose of study treatment will be used as the baseline in the analyses. If multiple

assessments are collected at the same date (and time if collected), the assessment with the latest database entry date (and time if collected) will considered as baseline.

Post baseline period:

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

6.2 Treatment Regimens

The treatment group "**as randomized"** will be retrieved from the IVRS system

- Arm A: Experimental arm: nivolumab + ipilimumab
- Arm B: Control arm: sunitinib

The treatment group "**as treated**" will be the same as the arm as randomized by IVRS. However, if a subject received the incorrect drug for **the entire period** of treatment, the subject's treatment group will be defined as the incorrect drug the subject actually received.

6.3 Populations for Analyses

- **All enrolled subjects**: All subjects who signed an informed consent form and were registered into the IVRS
- **All randomized subjects**: All subjects who were randomized to any treatment arm in the study. This population is considered as the secondary efficacy analysis population. Analysis of demography, protocol deviations, baseline characteristics, secondary efficacy analysis and outcome research analysis will be performed for this population.
- **Intermediate/poor risk subjects**: All randomized subjects with baseline IMDC prognostic score ≥ 1 at the time of randomization (IVRS). This is the primary efficacy analysis population. Analysis of demography, protocol deviations, baseline characteristics and primary efficacy analysis will be performed for this population.
- **All treated subjects**: All subjects who received any dose of study therapy. This is the primary dataset for drug exposure and safety analysis.

- **All treated intermediate/poor risk subjects**: All intermediate/poor risk subjects who received any dose of study therapy. **Favorable risk subjects**: All randomized subjects with baseline IMDC prognostic score = 0 at the time of randomization (IVRS). This population would be used for conducting exploratory analysis of efficacy endpoints.
- **PK subjects**: All subjects with available serum time-concentration data from randomized subjects dosed with nivolumab.
- **Immunogenicity subjects**: All subjects with available data from randomized subjects dosed with nivolumab.
- **PD-L1 treated subjects:** All subjects with a PD-L1 assessment at baseline who received any dose of study therapy.

All analyses will be performed using the treatment arm as randomized (intent to treat), with the exception of dosing and safety, for which the treatment arm as received will be used.

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, the bulleted titles in the following subsections describe tabulations of discrete variables, by the frequency and proportion of subjects falling into each category, grouped by treatment (with total). Percentages given in these tables will be rounded and, therefore, may not always sum to 100%. Continuous variables will be summarized by treatment group (with total) using the mean, standard deviation, median, minimum and maximum values. If a missing category is not being presented in the data display, only those subjects with non-missing values for the parameter being assessed are included in the percentage calculation.

Time to event distributions (i.e. progression free survival, overall survival, time to response, and duration of response) will be estimated using Kaplan-Meier techniques. When appropriate, the median along with 95% CI will be estimated using log-log transformation. Rates at fixed time points (e.g. OS at 12 months) will be derived from the Kaplan-Meier estimate along with their corresponding log-log transformed 95% confidence intervals.[7] Confidence intervals for binomial proportions will be derived using the Clopper-Pearson method.[8]

The unweighted difference in ORRs between the two treatment arms and corresponding asymptotic 95% CI will be estimated using a Newcombe method.[9]

Unless otherwise specified, the stratified log-rank test will be performed to test the comparison between time to event distributions (PFS and OS). Unless otherwise specified, the stratified hazard ratio between 2 groups along with CI will be obtained by fitting a stratified Cox model with the group variable as a unique covariate.

P-values from sensitivity analyses for efficacy endpoints are for descriptive purpose only and there will be no multiplicity adjustment for these analyses.

7.2 Study Conduct

7.2.1 Accrual

The accrual pattern will be summarized per country, investigational site, and per month for all enrolled, randomized and intermediate/poor risk subjects. Randomization date, first dosing date, country, investigational site will be presented in a by subject listing of accrual.

Furthermore, the accrual pattern will be summarized per month by the stratification factors.

7.2.2 Relevant Protocol Deviations

The following programmable deviations will be considered as relevant protocol deviations and summarized by treatment group and overall in all randomized subjects and in intermediate/poor risk subjects. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and non-programmable) eligibility and on-treatment protocol deviations will be reported through ClinSIGHT listings.

Eligibility:

- Subjects with baseline KPS < 70%
- Subjects who received prior systemic anti-cancer treatment in the metastatic setting
- Subjects without histologically confirmed RCC with a clear-cell component, documented advanced or metastatic (AJCC Stage IV) RCC

Eligibility (only for intermediate/poor risk subjects):

- Subjects with a baseline IMDC prognostic score < 1

On-study:

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

Listings will also be provided.

7.3 Study Population

Analyses in this section will be tabulated for all randomized subjects and for intermediate/poor risk subjects by treatment group as randomized, unless otherwise specified.

7.3.1 Subject Disposition

The total number of subjects enrolled (randomized or not randomized) will be presented along with the reason for not being randomized. This analysis will be performed only on the all enrolled subjects population only.

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

Number of subjects randomized but not treated along with the reason will be tabulated by treatment group as randomized.

A subject listing for all randomized subjects will be provided showing the subject's randomization date, first and last dosing date, off study date and reason for going off-study. A subject listing for subjects not randomized will also be provided, showing the subject's race, gender, age, consent date and reason for not being randomized.

7.3.2 Demographics and Other Baseline Disease Characteristics

The following baseline characteristics will be summarized by treatment arm as randomized:

- Age
- Age categorization (< 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85, ≥ 75, ≥ 65)
- Gender (Male vs. Female)
- Race (White, Black or African American, Asian, Other)
- Ethnicity (Hispanic/Latino and Not Hispanic/Latino)
- Karnofsky performance status(70, 80, 90, 100)
- Baseline IMDC prognostic score (0, 1-2, ≥ 3) (source: CRF)
- Prior nephrectomy
- Prior radiotherapy
- Time from initial disease diagnosis to randomization (<1 year, ≥1 year)
- LDH level (≤ 1.5 x ULN, >1.5 x ULN)
- Hemoglobin (<LLN, ≥ LLN)
- Corrected Calcium (≤ 10 mg/dl, >10mg/dl)
- Alkaline phosphatase (< ULN, ≥ ULN)
- Region (per IVRS)
- Baseline PD-L1+ status based on a 1% cut off
- Baseline PD-L1+ status based on a 5% cut off
- Baseline PD-L1+ status based on a 10% cut off
- Sites of diseases (all lesions)
- Number of disease sites per subject (all lesions)
- Tumor burden: sum of the diameters of target lesions at baseline
- Pre-treatment events: summarized by worst CTC grade presented by SOC/PT

7.3.3 Medical history

General medical history will be tabulated and also listed by subject.

7.3.4 *Prior* therapy agents

- Prior adjuvant or neo-adjuvant therapy agents for localized or locally advanced RCC will be summarized.

7.3.5 Baseline examinations

Subjects with abnormal baseline physical examination will be tabulated by examination criteria and by treatment arm.

7.3.6 Baseline Physical Examination

Summary of baseline height and weight will be tabulated and presented.

7.3.7 Discrepancies Between IVRS and CRF stratification factors

Summary tables (cross-tabulations) by treatment arm for each baseline stratification factor will be provided to show any discrepancies between what was reported through IVRS vs. CRF data.

- Baseline IMDC prognostic score (0 vs. 1-2 vs. ≥ 3)

7.4 Extent of Exposure

Listings will include all available exposure data. Analyses will be performed by treatment group "as treated" in all treated subjects, unless otherwise specified.

7.4.1 Administration of Study Therapy

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

- Time from randomization to first dose of study therapy (0 to 3 days, > 3 to 7, > 7 to 14, > 14 to 21, > 21 to 28, > 28)

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

- Number of doses received (nivolumab, ipilimumab, sunitinib):
- Cumulative dose (nivolumab, ipilimumab, sunitinib)
- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%. (nivolumab, ipilimumab, sunitinib)

Duration of treatment will be presented by treatment group using a Kaplan-Meier curve whereby the last dose date will be the event date for those subjects who are off study therapy. Median duration of treatment and associated 95% CI will be provided. Subjects who are still on study therapy will be censored on their last dose date.

A by-subject listing of dosing of study medication (record of study medication, infusion details, and dose changes) and a listing of batch numbers will be also provided.

[redacted]

[redacted]

[redacted]

7.5 Efficacy

The formal analysis of ORR in the intermediate and poor-risk randomized subjects will be performed when these patients have an at least 6 month minimum follow-up from the completion of enrollment. An administrative allocation of 0.001 alpha will be used. At the time the ORR analysis, a reduced set of the total analyses defined in this document will be performed (e.g. no PFS/OS analysis). The relevant analyses will be specifically defined in the data presentation plan.

Principal analyses of PFS and ORR will be based on the IRRC evaluation, unless noted otherwise. Unless stated otherwise, whenever a stratified analysis is specified using intermediate/poor risk subjects, the following stratifications factors (recorded at randomization as per IVRS) will be used:

- IMDC prognostic risk score (1-2 vs. 3-6)
- Region (US vs.Canada/W.Europe/N.Europe vs. ROW)

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

- IMDC prognostic risk score (0 vs.1-2 vs. 3-6)
- Region (US vs.Canada/W.Europe/N.Europe vs. ROW)

For assessing the secondary objectives of this study, a hierarchical testing procedure[10] will be used so that the overall experiment-wise Type I error rate is 0.05.

The key secondary objectives among all randomized subjects will be tested after conducting the primary objective analyses on intermediate/poor risk subjects.

The formal testing of ORR per IRRC using 95% exact CIs among all randomized subjects will take place if the 95% exact CI of ORR per IRRC across treatment groups among all randomized intermediate/poor risk subjects do not overlap. The analysis of ORR is descriptive so no p-values for rate differences will be reported.

Similarly, the formal testing of PFS as per IRRC, at a two sided 0.009 significance level, among all randomized subjects will take place if PFS per IRRC among intermediate/poor risk subjects is statistically significant. Likewise, the testing of OS, at a two sided 0.04 significance level, among all randomized subjects will take place only if OS intermediate/poor risk subjects is statistically significant.

All p-values reported will be two-sided. Confidence Interval for co-primary and secondary endpoint analyses included in hierarchy (PFS and OS) will be based on nominal significance level adjusted for co-primary endpoints and interim analyses to preserve overall type one error rate (See sections 7.5.1, 7.5.6, 7.5.7 and 7.5.8 for details). Alpha (α) for the confidence interval will be the same as nominal significance level for hypothesis testing. CIs for other endpoints will be at the two-sided 95% level. The p-values presented in the clinical study report will be rounded to the fourth decimal place. Point estimates and confidence bounds for efficacy variables will be rounded to the second decimal place.

7.5.1 Analysis of Objective Response Rate towards Primary Objective

One of the primary objectives of the study is to estimate the objective response rate per IRRC in the two treatment arms among intermediate and poor risk subjects. For the ORR per IRRC analysis both the final and interim CRF pages collect BOR following this algorithm:

- Apply final BOR if available (from level1 VA), otherwise use the most recent interim BOR (from level1 TM)

Estimates of response rate, along with its exact two-sided 95% CI by Clopper-Pearson method, will be computed within each treatment arm. A two sided 95% CI for difference of response rate between the treatment arms will also be computed.

Sensitivity analysis based on investigator-determined ORR may also be performed. DOR and TTR will also be evaluated. At the time of the formal ORR analysis, there will be no formal analysis of PFS and OS.

One of the secondary objectives of the study is to estimate the objective response rate in the two treatment arms among all randomized subjects.

The number and percentage of subjects in each category of best overall response per IRRC (complete response [CR], partial response [PR]], stable disease [SD], progressive disease [PD], or unable to determine [UD]) according to the IRRC will be presented, by treatment group. An estimate of the response rate and an associated exact two-sided 95% CI (Clopper and Pearson[11]) will be presented, by treatment group.

A 2-sided, 95% confidence interval for the difference of ORR between treatment arms will be computed for all randomized subjects by the method of DerSimonian and Laird[12], using a fixed-effects model (setting Δ^2 equal to zero), adjusting for the stratification factors. The weighted response rate difference and 95% CI can be determined using the following formula:

$$\hat{\theta} = \frac{\sum_{i=1}^{12} \hat{\theta_i} w_i}{\sum_{i=1}^{12} w_i} \sim N(\theta, 1/\sum_{i=1}^{12} w_i)$$

where $\hat{\theta_i}$ is the response rate difference of the i[th] stratum and $w_i = 1/var(\hat{\theta_i})$.

Similar analyses will be repeated based on the investigator's assessment of ORR. A cross tabulation of IRRC best response versus the investigator best response will be presented, by treatment group.

7.5.2 Analysis of Progression-Free Survival towards Primary Objective

One of the primary objectives of the study is to compare the progression-free survival (as determined by IRRC) of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC. All the analyses outlined in this section are specified for intermediate/poor risk subjects population.

The principal analysis of PFS (as determined by IRRC) will be to compare the two treatment arms via two sided 0.009 stratified log-rank test among intermediate/poor risk subjects. The primary definition of PFS will be used in this analysis. The two-sided log-rank p-value will be reported.

The estimate of the PFS hazard ratio, of nivolumab combined with ipilimumab to sunitinib monotherapy, will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. Ties will be handled using the Exact method. A two-sided, 99.1% CI for the hazard ratio will also be presented.

The PFS function for each treatment arm will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each arm will be computed via the log-log transformation method. Estimates for one-year and two-year PFS rates will be presented along with their associated 95% CIs. Minimum follow-up must be greater than or equal to the time-point to generate the rate. These estimates will come from the KM curve and their standard errors (SEs) and associated CIs, will be computed using log-log transformed 95% confidence intervals[13].

The method of Gail and Simon[14] will be used to test for a qualitative interaction between treatment and strata, IMDC prognostic risk score (1-2 vs. 3-6) and Region (US vs. Canada/W.Europe/N.Europe vs. ROW). This test will be conducted at α= 0.10 level. The p-value reported from this specific analysis is for descriptive purposes only.

The proportional hazards assumption will be assessed via the following hazard rate model, which contains a time dependent covariate:

$$\lambda(t, z) = \lambda_i(t) e^{(\{b_1 + b_2 \times [\log(t)]\} \times Z)}, \qquad i = \{1 - 6\}$$

where i=1-6 corresponds to each of the six levels the stratum can take, and Z is the treatment indicator, which is equal to 1 for the combination arm and 0 for the control arm.. The null hypothesis, that the proportional hazards assumption is valid, i.e., that $b_2=0$, will be tested against the alternative hypothesis that $b_2 \neq 0$ using a Wald statistic at $\alpha= 0.10$ level. The p-value reported from this specific analysis is for descriptive purposes alone. A plot of smoothed scaled Schoenfeld residuals of the above model will be used to graphically illustrate the evolution of the hazard ratio over time.

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

Analyses of PFS will also be conducted based on the ITT definition (secondary definition) of PFS. These analyses will be the same as those specified above.

The status of subjects who are censored (as per primary definition of PFS) in the PFS KM analysis will be tabulated for each randomized treatment group using the following categories:

- On-study (on-treatment, in follow-up)
- Off-study (lost to follow-up, withdrawn consent, never treated)
- Received subsequent anticancer therapy

7.5.3 Supportive Analyses of Progression-Free Survival

The following sensitivity analyses will be conducted using both the primary and the ITT definitions of PFS in the intermediate/poor risk subjects:

1. Delayed effect of immunotherapy interventions may cause a late separation in the progression free survival KM curves and non-proportional hazards as was observed in the second line phase 3 mRCC study (CA209-025). The principal analysis of PFS (as determined by IRRC) will be to compare the two treatment arms via two sided 0.009 stratified weighted log-rank test among intermediate/poor risk subjects. The primary definition of PFS will be used in this analysis. The two-sided stratified weighted log-rank p-value will be reported using G (rho = 0, gamma = 1) weights, in the terminology of Fleming and Harrington[15].

   The Fleming Harrington test can be unstable, so it is possible, though uncommon, that the p-value for this trial will not be estimable. In this case, the primary analysis will default to using the two sided 0.009 stratified log-rank test among intermediate/poor risk subjects (as specified in section 7.5.2).

   The estimate of the PFS hazard ratio in the period following 6 months, of nivolumab combined with ipilimumab compared to sunitinib monotherapy, will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. In this model, period is a binary variable indicating pre- vs post- 6 months. The second line phase 3 mRCC study (CA209-025) served as the basis for the 6 month delayed treatment effect in PFS. Ties will be handled using the exact method. A two-sided 99.1% CI for the hazard ratio will also be presented.

2. A multivariate Cox regression model will be used in order to estimate the treatment effect after adjustment for possible imbalances in known or potential prognostic factors. The factors used in the randomization, which, by definition, will be balanced across arms, will still be included in the model as stratification factors. However, all additional factors will be incorporated as covariates. The additional factors, which are all measured at baseline, will include:

    a. LDH (≤ 1.5 x ULN, > 1.5 x ULN)

    b. Previous Nephrectomy (Yes, No)

The level of the covariate normally associated with the worst prognosis will be coded as the reference level:

The hazard ratio associated with treatment and with each of the baseline covariates will be presented along with associated 99.1% CIs.

3. PFS using stratification factors as obtained from the baseline CRF pages (instead of IVRS). The hazard ratio associated with treatment will be presented along with the associated two-sided 99.1% CIs. This analysis will be performed only if at least one stratification variable/factor at randomization (as per IVRS) and baseline are not concordant for at least 10% of the randomized intermediate / poor risk subjects.
4. PFS using the investigator's assessment. The hazard ratio associated with treatment and median PFS will be presented along with the associated two-sided 99.1% CIs.
5. PFS using an un-stratified log rank test. The hazard ratio associated with treatment will be presented along with the associated two-sided 99.1% CIs.
6. PFS using an unstratified Cox proportional hazards model, adjusted, using as covariates only the two stratification factors used in randomization. The hazard ratio associated with treatment will be presented along with the associated two-sided 99.1% CIs.
7. PFS for subjects with no relevant deviation. This analysis will be conducted only if there are more than 10% subjects with relevant protocol deviations. The hazard ratio associated with treatment will be presented along with the associated two-sided 99.1% CIs.

A by-subject listing will be presented including treatment arm, PFS duration under the primary definition, PFS duration on the ITT definition, whether the subject was censored under the primary definition, and if censored, the reason, and whether the subject was censored under the ITT definition, and if censored, the reason.

7.5.4 Concordance Between IRRC and Investigator Assessments of Progression

For the purpose of assessing concordance between the IRRC and investigator tumor assessments among intermediate/poor risk subjects, progression status will be categorized as documented progression, death or censored. A cross tabulation between the IRRC and the investigator

progression status will be presented, by treatment group. The secondary definition of PFS (ITT) will be used for this analysis.

The number of subjects with the same timing of IRRC and investigator documented progression will be summarized. In addition, a frequency table of the time between the date of the IRRC documented progression and the date of the investigator documented progression (weeks) will be presented by treatment group, for subjects who had documented progression according to both the IRRC and the investigator at different time points. The time between dates will be defined as:

$$\frac{\left(\begin{array}{c} date\ of\ IRC\ documented\ progression - \\ date\ of\ investigator\ documented\ progression \end{array}\right)}{7}$$

The time between dates will be categorized as < -12 weeks, -12 weeks to < -8 weeks, -8 weeks to < -4 weeks, -4 weeks to < -2 weeks, -2 weeks to < 0 weeks, 0 weeks to < 2 weeks, 2 weeks to < 4 weeks, 4 weeks to < 8 weeks, 8 weeks to < 12 weeks, ≥ 12 weeks. Subjects who only progressed per investigator and had a death event per IRRC will also be summarized along with the difference in the timing of events. The time between dates will be categorized as < 10 weeks, 10 to < 20 weeks, ≥ 20 weeks.

A by subject listing of IRRC and investigator PFS status and the time between progression dates according to the IRRC and the investigator will be provided.

7.5.5 Subset Analyses of Progression-Free Survival

The influence of baseline and demographic characteristics on the treatment effect among intermediate/poor risk subjects will be explored via exploratory subset analyses for the following factors

- Age categorization (< 65 vs. ≥ 65 - < 75 vs. ≥ 75)
- Gender (Male vs. Female)
- Race
- Region (US vs.Canada/W.Europe/N.Europe vs. ROW)
- Karnofsky performance status(< 90 vs. ≥ 90)
- Baseline IMDC prognostic score (1-2, ≥ 3) (source: CRF)
- Prior adjuvant or neo-adjuvant therapy for localized or locally advanced RCC (Yes, No)
- Prior Nephrectomy (Yes, No)
- Prior Radiotherapy (Yes, No)
- Time from initial disease diagnosis to randomization (< 1 year, ≥ 1 year)
- LDH level (≤ 1.5 x ULN , > 1.5 x ULN)
- Hemoglobin (< LLN , ≥ LLN)
- Corrected Calcium (≤ 10 mg/dl, > 10mg/dl)
- Alkaline phosphatase (< ULN, ≥ ULN)
- Baseline PD-L1+ status based on a 1% cut off
- Baseline PD-L1+ status based on a 5% cut off

- Baseline PD-L1+ status based on a 10% cut off

A forest plot of the PFS hazard ratios (along with the 95% CIs) will be produced for each level of the subgroups listed above.

All the above mentioned analyses (except age, race, region, and gender) will be conducted if the number of subjects in each subgroup is more than 20. Estimates of median PFS would be computed for all the subsets.

7.5.6 Analysis of Overall Survival towards the Primary Objective

One of the primary objectives of the study is to compare the overall survival of nivolumab combined with ipilimumab to sunitinib monotherapy in intermediate and poor-risk subjects with previously untreated mRCC. All the analyses outlined in this section are specified for intermediate/poor risk subjects population.

Overall survival will be compared between the treatment arms at the interim and final analyses, using stratified log-rank test. The stratification factors will be those used in the analysis of PFS. An O'Brien and Fleming α-spending function will be employed to determine the nominal significance levels for the interim and final analyses. The stratified hazard ratio between the treatment groups will be presented along with 100*(1- α)% CI (adjusted for interim). In addition, two-sided p-value will also be reported for the primary analysis of OS.

All analyses performed for PFS (detailed in section 7.5.1) will be repeated for OS. Supportive analyses 1, 2, 4, 5 and 6 of PFS (detailed in section 7.5.3) as well as the subset analyses (detailed in section 7.5.5) will also be repeated for OS.

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

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

Estimates for 1, 2 and 3-year OS rates will be presented along with their associated 95% CIs. These analyses will be only performed if the minimum follow-up for OS has reached corresponding to that endpoint. These estimates and their standard errors (SEs) will be come from the KM curve for use in constructing CIs computed using log-log transformed 95% confidence intervals.

7.5.7 Analysis of Progression-Free Survival towards the Secondary Objective

One of the key secondary objectives of the study is to compare the progression-free survival (as determined by IRRC) of nivolumab combined with ipilimumab to sunitinib monotherapy in all randomized subjects with previously untreated mRCC.

If the formal comparison of PFS (as per IRRC) in the intermediate/poor risk subjects is found to be statistically significant, then PFS (as determined by IRRC) will be compared between the two treatment arms via two sided 0.009 stratified log-rank test among intermediate/poor risk subjects.

The primary definition of PFS will be used in this analysis. The two-sided log-rank p-value will be reported. Further analysis of PFS will include estimation of the hazard ratio and estimation of the PFS distribution in each treatment group.

The estimate of the PFS hazard ratio, of nivolumab combined with ipilimumab to sunitinib monotherapy, will be calculated using a stratified Cox proportional hazards model, with treatment as the sole covariate. Ties will be handled using the Exact method. A two-sided, 99.1% CI for the hazard ratio will also be presented.

The PFS function for each treatment arm will be estimated using the KM product limit method and will be displayed graphically. A two-sided 95% CI for median PFS in each arm will be computed via the log-log transformation method. Estimates for one-year and two-year PFS rates will be presented along with their associated 95% CIs. Minimum follow-up must be greater than or equal to the time-point to generate the rate. These estimates and their standard errors (SEs) will be come from the KM curve for use in constructing CIs computed using log-log transformed 95% confidence intervals.

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

Analyses of PFS will also be conducted based on the ITT definition of PFS. These analyses will be the same as those specified above.

The status of subjects who are censored (as per primary definition of PFS) in the PFS KM analysis will be tabulated for each randomized treatment group using the following categories:

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

Received subsequent anticancer therapy

Supportive analyis (3) of PFS (detailed in section 7.5.3) as well as the subset analyses (detailed in section 7.5.5) will also be repeated for PFS among all randomized subjects, if the principal comparison of PFS among all randomized subjects is found to be statistically significant.

7.5.8 Analysis of Overall Survival towards the Secondary Objective

One of the key secondary objectives of the study is to compare the overall survival of nivolumab combined with ipilimumab to sunitinib monotherapy in all randomized subjects with previously untreated mRCC. The testing of OS, at a two sided 100*(1-α)% significance level, among all randomized subjects will take place only if OS intermediate/poor risk subjects is statistically significant.

Analyses of OS towards secondary objective will be similar to PFS analyses outlined in 7.5.7.

7.5.9 Current status of PFS and OS follow-up

Time from last tumor assessment to data cut-off in months will be summarized by treatment arm and overall for all randomized subjects. Subjects who have a PFS event will be considered as current for this analysis. The ITT definition of PFS will be used for this summary.

In addition Kaplan-Meier plots of time from randomization to post-baseline tumor assessment will be produced by treatment arm for the first twelve assessments.

Current status of OS follow-up will be summarized in months, by computing the time from "last known alive" date to data cut-off date. Subjects who have a death event will be considered as current for this analysis.

By-subject listings will also be produced to accompany the subject time from last tumor assessment table.

7.5.10 Analysis of Objective Response

One of the secondary objectives of the study is to estimate the objective response rate in the two treatment arms among intermediate/poor risk and all randomized subjects separately.

The number and percentage of subjects in each category of best overall response per IRRC (complete response [CR], partial response [PR]], stable disease [SD] (including Non-CR/Non-PD), progressive disease [PD], or unable to determine [UD]) according to the IRRC will be presented, by treatment group. An estimate of the response rate and an associated exact two-sided 95% CI (Clopper and Pearson[16]) will be presented, by treatment group.

A 2-sided, 95% confidence interval for the difference of ORR between treatment arms will be computed for all randomized subjects by the method of DerSimonian and Laird[17], using a fixed-effects model (setting Δ^2 equal to zero), adjusting for the stratification factors. The weighted response rate difference and 95% CI can be determined using the following formula:

$$\hat{\theta} = \frac{\sum_{i=1}^{12} \hat{\theta}_i w_i}{\sum_{i=1}^{12} w_i} \sim N(\theta, 1/\sum_{i=1}^{12} w_i)$$

where $\hat{\theta}_i$ is the response rate difference of the ith stratum and $w_i = 1/var(\hat{\theta}_i)$.

Similar analyses will be repeated based on the investigator's assessment of ORR. A cross tabulation of IRRC best response versus the investigator best response will be presented, by treatment group.

7.5.11 Subset Analyses of Objective Response

The influence of baseline and demographic characteristics on the treatment effect will be explored via exploratory subset analysis. The subsets will be the same as those analyzed for PFS and will be reported based on the IRRC assessment of ORR.

The un-weighted differences in ORR between the two treatment groups and corresponding 95% two-sided CI using the method of Newcombe will be provided.

7.5.12 Time to Tumor Response, Time in Response, and Duration of Response

The distributions of duration of response will be estimated, by arm, using the KM product limit method. The KM estimates will be presented graphically and tables will be produced presenting number of events, number of subjects involved, medians, and 95% CIs for the medians.

Time to tumor response, which does not involve censoring, will be summarized by treatment group, using descriptive statistics.

A by-subject listing will be presented including treatment arm, time in tumor response, whether subject was censored for time in tumor response, and if so, the reason, duration of response, whether the subject was censored for duration of response, and, if so, the reason.

7.5.13 Interim Analysis of Overall Survival

An independent statistician external to BMS will perform the analysis. In addition to the formal planned interim analysis for OS, the Data Monitoring Committee (DMC) will have access to periodic un-blinded interim reports of efficacy and safety to allow a risk/benefit assessment. Details are included in the DMC charter.

Two interim analyses of OS are planned. The first and second interim analyses are scheduled at the time of final PFS analysis when 370 deaths (approximately 58% of the targeted OS events) are expected and after 479 deaths (approximately 75% of total deaths) have been observed, respectively, among intermediate/poor risk subjects based on above accrual rate and the exponential distribution in each arm. These formal comparisons of OS will allow for early stopping for superiority, and the boundaries for declaring superiority will be derived based on the actual number of deaths using Lan-DeMets spending function with O'Brien and Fleming type of boundary in EAST v5.4. If the first interim analysis is performed exactly at 370 deaths, the boundary in terms of statistical significance for declaring superiority would be 0.0045 (HR=0.74, 6.9 months improvement in median OS) and if the second interim analysis is performed at exactly 479 deaths, the boundary in terms of statistical significance at the interim analysis for declaring superiority would be 0.0131 (or 0.8 with regard to HR boundary, which corresponds to 5.1 months improvement in median OS under the assumed control arm hazard function). The boundary for declaring superiority in terms of statistical significance for the final analysis after 639 events would be 0.0354.

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

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

Implications of OS Interim Analysis

At the time of the formal interim analysis for superiority of OS, the DMC may recommend continuing or stopping the trial. If the trial continues beyond the interim look, the nominal critical point for the final OS analysis will be determined using the recalculated information fraction at the time of the interim analysis, as described above. The final OS hazard ratio and corresponding confidence interval will be reported whereby the confidence interval will be adjusted accordingly (i.e. using the recalculated nominal α level at the final analysis).

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

7.6 Safety

Safety summary tables will be generated for all treated subjects. Listings will include all available data.

7.6.1 All Adverse Events

See CORE Safety SAP[2]. In addition, summary tables will be presented for all treated intermediate/poor risk subjects.

7.6.2 Deaths

See CORE Safety SAP[2]. In addition, summary tables will be presented for all treated intermediate/poor risk subjects.

7.6.3 Serious Adverse Events

See CORE Safety SAP[2]. In addition, summary tables will be presented for all treated intermediate/poor risk subjects.

7.6.4 Adverse Events Leading to Discontinuation/Modification of Study Therapy

See CORE Safety SAP[2].

7.6.5 Multiple Events

See CORE Safety SAP[2].

7.6.6 Other Observations Related to Safety

See CORE Safety SAP[2].

7.6.7 Select Adverse Events

See CORE Safety SAP[2].

7.6.8 Immune-Mediated Adverse Events

See CORE Safety SAP[2].

7.6.9 Other Events of Special Interest

See CORE Safety SAP[2].

7.6.10 Clinical Laboratory Evaluations

7.6.10.1 Hematology

See CORE Safety SAP[2]. In addition, summary tables will be presented for all treated intermediate/poor risk subjects.

7.6.10.2 Serum Chemistry

Amylase and lipase will be summarized in addition to the serum chemistry parameters described in the CORE safety SAP[2].

7.6.11 Immunogenicity

See CORE Safety SAP[2].

7.6.12 Vital Signs and Physical Findings

See CORE Safety SAP[2].

[redacted]

[redacted]

[redacted]

8 CONVENTIONS

The following conventions may be used for imputing partial dates for analyses requiring dates:

For missing and partial adverse event onset dates, imputation will be performed using the Adverse Event Domain Requirements Specification[18]. Missing and partial Non-Study Medication Domain dates will be imputed using the derivation algorithm described in 4.3.3 of BMS Non-Study Medication Domain Requirements Specification[19].

For death dates, the following conventions will be used for imputing partial dates:

- If only the day of the month is missing, the 1st of the month will be used to replace the missing day. The imputed date will be compared to the last known date alive and the maximum will be considered as the death date.
- If the month or the year is missing, the death date will be imputed as the last known date alive
- If the date is completely missing but the reason for death is present the death date will be imputed as the last known date alive

For date of progression, the following conventions will be used for imputing partial dates:

- If only the day of the month is missing, the 1st of the month will be used to replace the missing day*.
- If the day and month are missing or a date is completely missing, it will be considered as missing.

*In cases where the date of death is present and complete, the imputed progression date will be compared to the date of death. The minimum of the imputed progression date and date of death will be considered as the date of progression.

For other partial/missing dates, the following conventions may be used:

- If only the day of the month is missing, the 15th of the month will be used to replace the missing day.
- If both the day and the month are missing, "July 1" will be used to replace the missing information.
- If a date is completely missing, it will be considered as missing.

The following conversion factors will be used to convert days to months or years: 1 month = 30.4375 days and 1 year = 365.25 days.

Duration (e.g. time from first diagnosis to first dosing date, duration of response, and time to response) will be calculated as follows:

Duration = (Last date - first date + 1)

All statistical analyses will be carried out using SAS (Statistical Analysis System software, SAS Institute, North Carolina, USA) unless otherwise noted.

9 CONTENT OF REPORTS

All analyses described in this SAP will be included in the Clinical Study Report(s) except where otherwise noted. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

[redacted]

[redacted]