Official Title of Study:

Randomized, Open-Label, Phase 3 Trial of Nivolumab plus Ipilimumab or Nivolumab plus Platinum-Doublet Chemotherapy versus Platinum-Doublet Chemotherapy in Early Stage NSCLC
(CheckMate 816: CHECKpoint pathway and nivoluMAb clinical Trial Evaluation 816)

PROTOCOL(S) CA209-816

NCT Number: NCT02998528

Document Date (Date in which document was last revised): August 11, 2021

STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT

RANDOMIZED, OPEN-LABEL, PHASE 3 TRIAL OF NIVOLUMAB PLUS IPILIMUMAB OR NIVOLUMAB PLUS PLATINUM-DOUBLET CHEMOTHERAPY VERSUS PLATINUM-DOUBLET CHEMOTHERAPY IN EARLY STAGE NSCLC

PROTOCOL(S) CA209816

VERSION # 3.0

DATE: 11-AUG-2021

TABLE OF CONTENTS

[redacted]

[redacted]

LIST OF TABLES

LIST OF FIGURES

[redacted]

1 BACKGROUND AND RATIONALE

Approximately 80% of lung cancer cases are non-small cell lung cancer (NSCLC), with most patients presenting with late-stage disease. At initial diagnosis, 20% of patients present with stage I or II disease, whereas 30% present with stage III disease and 50% with stage IV disease. With enhanced lung cancer screening techniques, the percentage of patients diagnosed during the early stages may increase over the duration of the trial. A standard TNM staging system is used to determine the staging for NSCLC. Patients with pathologic stage I NSCLC have a 5-year survival of approximately 60%. Stage II to III NSCLC patients have a 5-year survival of approximately 25% to 40%.[1] Surgical resection remains the mainstay of treatment for stage I and II patients; however, despite potentially curative surgery, approximately 50% of stage IB and 60-75% of stage II NSCLC patients will relapse and eventually die of their disease.[2,3] A rational approach to improve survival in these patients is to eradicate micrometastatic disease and to minimize the risk of relapse after adjuvant or neoadjuvant chemotherapy.

The phase 3 study, CA209816, will evaluate the clinical efficacy and will establish the safety of nivolumab plus platinum doublet chemotherapy and nivolumab plus ipilimumab, in resectable lung cancer. Specifically, this study will compare EFS and pCR rate among participants treated with neoadjuvant nivolumab plus platinum doublet chemotherapy vs participants treated with platinum doublet chemotherapy, and will describe pCR rate and EFS for those treated with neoadjuvant nivolumab plus ipilimumab in Stage Ib-IIIa NSCLC.

This document contains description of the statistical analyses that will be conducted for the clinical study report (CSR) of study CA209816.

Research Hypothesis:

In participants with stage IB (≥ 4 cm), II or IIIA (N2) NSCLC considered resectable by the local multidisciplinary team, administration of neoadjuvant nivolumab plus platinum doublet chemotherapy (up to 3 cycles) has superior efficacy to neoadjuvant platinum doublet chemotherapy (up to 3 cycles).

Schedule of Analyses:

Formal analysis of Pathological Complete Response (pCR) will occur after the 350 randomized participants on Arms B and C from start of 1:1:1 randomization have an opportunity for surgery and is projected to occur approximately 30 months after 1:1:1 randomization.

Two formal interim analyses for Event Free Survival (EFS) are planned after 148 and 167 events have been observed on Arms B and C after start of 1:1:1 randomization, respectively. The second interim analysis may take place one year after the first interim analysis in case the required number of events is not yet reached at that time. This is projected to occur approximately 48 and 58 (or max 60) months after start of 1:1:1 randomization. The formal interim comparisons of EFS will allow for determination of superiority. If the study continues beyond these interim analyses, the final analysis will be conducted after approximately 185 EFS events have been observed on Arms B and C from start of 1:1:1 randomization, or at a maximum 4 years after the last subject's randomization (December 2023).

[redacted]

2 STUDY DESCRIPTION

2.1 Study Design

This is an open-label, randomized clinical trial of up to 3 cycles of neoadjuvant nivolumab (3 mg/kg every 2 weeks) and a single dose of 1 mg/kg dose of ipilimumab, nivolumab 360mg flat dose plus platinum doublet chemotherapy (up to 3 cycles), or platinum doublet chemotherapy (up to 3 cycles) as neoadjuvant treatment in participants with early stage (Stage IB [≥ 4 cm], II, and resectable IIIA [N2]) NSCLC.

The original study design (before revised protocol 02) had two arms. After signing the informed consent form and upon confirmation of the participant's eligibility, participants were randomized in an open-label fashion (1:1 ratio) to either neoadjuvant nivolumab plus ipilimumab or platinum doublet chemotherapy.

Revised protocol 02 added a new, neoadjuvant nivolumab plus platinum doublet chemotherapy arm. When the third arm had opened and as each site had received IRB/EC approval of revised protocol 02, the interactive response technology IRT switched to a 1:1:1 randomization at the respective site. Starting from that point on, the sites were only enrolling under revised protocol 02.

Revised protocol 03 withholds randomization into the arm of neoadjuvant nivolumab plus ipilimumab but continues randomizing eligible participants into either neoadjuvant nivolumab plus platinum doublet chemotherapy arm or platinum doublet chemotherapy arm. Participants already randomized in the original 2-arm part (neoadjuvant nivolumab plus ipilimumab vs neoadjuvant chemotherapy) and in the arm of neoadjuvant nivolumab plus ipilimumab in 3-arm part defined by revised protocol 02 will remain in trial and continue scheduled trial procedures. The primary population for comparisons of the primary endpoints is the subjects concurrently randomized in arms B and C (as of revised protocol 02).

As of Revised protocol 03, participants will be randomized between 2 arms in a 1:1 ratio to neoadjuvant nivolumab plus platinum doublet chemotherapy or platinum doublet chemotherapy. Eligible participants will be stratified by:

- PD-L1 expression (≥1% or <1%/not evaluable/indeterminate)
- Disease stage (IB/II vs IIIA)
- Gender

The treatment arms are as follows:

Arm A treatment: Participants randomized into Arm A received nivolumab 3 mg/kg IV over 30 minutes every 2 weeks for up to 3 doses (ie, 6 weeks of treatment; each cycle is 14 days). With Cycle 1 only, nivolumab was followed by a single dose ipilimumab 1 mg/kg IV over 30 minutes.

Arm B treatment: Participants randomized into Arm B will receive investigator-choice platinum doublet chemotherapy in 3-week cycles up to a maximum of 3 cycles (ie, 9 weeks of treatment; each cycle is 21 days):

- Regimen 1:
  - Vinorelbine 25 mg/m^2 or 30 mg/m^2 IV (per local prescribing information) push over 10 minutes or per institutional standard on Days 1 and 8
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following vinorelbine
- Regimen 2:
  - Docetaxel 60 mg/m^2 or 75 mg/m^2 IV (per local prescribing information) over 60 minutes or per institutional standard on Day 1
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following docetaxel
- Regimen 3 (squamous histology):
  - Gemcitabine 1000 mg/m^2 or 1250 mg/m^2 (per local prescribing information) IV over 30 minutes or per institutional standard on Days 1 and 8
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following gemcitabine
- Regimen 4 (non-squamous histology only):
  - Pemetrexed 500 mg/m^2 IV over 10 minutes or per institutional standard on Day 1
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following pemetrexed
- Regimen 5:
  - Paclitaxel 175 or 200 mg/m^2 IV over 180 minutes or per institutional standard on Day 1
  - Carboplatin AUC 5 or 6 IV over 30 minutes or per institutional standard on Day 1, immediately following paclitaxel

Arm C treatment: Participants randomized into Arm C will receive nivolumab 360 mg IV plus platinum doublet chemotherapy in 3-week cycles up to a maximum of 3 cycles of chemotherapy (ie, 9 weeks of treatment; each cycle is 21 days)

- Non-squamous NSCLC: nivolumab at a flat dose of 360 mg as 30-minute IV infusion on Day 1, followed by pemetrexed at a dose of 500 mg/m^2 IV over 10 minutes or per institutional standard and cisplatin at a dose of 75 mg/m^2 IV over 120 minutes or per institutional standard of a 3-week treatment cycle, for up to 3 cycles.
- Squamous NSCLC: nivolumab at a dose of flat dose of 360 mg as 30 minute IV infusion on Day 1, followed by gemcitabine at a dose of 1000 mg/m^2 or 1250 mg/m^2 (per local prescribing information) for a 30 minute IV infusion or per institutional standard and cisplatin at a dose of 75 mg/m^2 as a 120-minute IV infusion or per institutional standard, of a 3-week treatment cycle for up to 3 cycles. Gemcitabine will also be administered at a dose of 1000 mg/m^2 or 1250 mg/m^2 as a 30 minute IV infusion or per institutional standard on day 8 of each 3-week treatment cycle.

- Any histology: nivolumab at a flat dose of 360 mg as 30-minute IV infusion on Day 1, followed by paclitaxel 175 or 200 mg/m^2 IV over 180 minutes or per institutional standard and carboplatin AUC 5 or 6 IV over 30 minutes or per institutional standard of a 3-week treatment cycle, for up to 3 cycles.

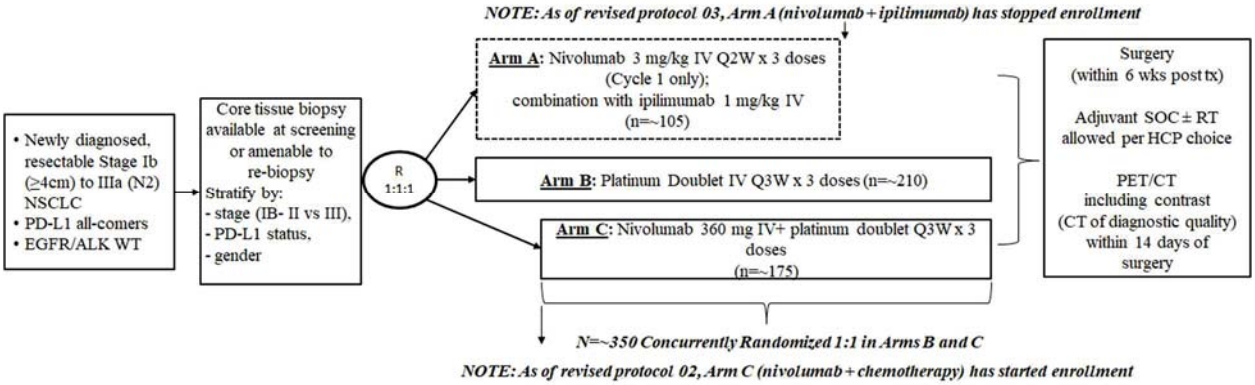
Following the completion of neoadjuvant treatment, all participants who remain operative candidates will undergo definitive surgery for their NSCLC within 6 weeks after completing neoadjuvant treatment.

Following definitive surgery, participants in each arm may receive up to 4 cycles of adjuvant chemotherapy with or without radiation per institutional standard at the discretion of the investigator. Investigators may choose from the following post-operative regimens:

- Regimen 1:
  - Vinorelbine 25 mg/m^2 or 30 mg/m^2 IV (per local prescribing information) push over 10 minutes or per institutional standard on Days 1 and 8
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following vinorelbine
- Regimen 2:
  - Docetaxel 60 mg/m^2 or 75 mg/m^2 IV (per local prescribing information) over 60 minutes or per institutional standard on Day 1
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following docetaxel
- Regimen 3 (squamous histology):
  - Gemcitabine 1000 mg/m^2 or 1250 mg/m^2 IV (per local prescribing information) over 30 minutes or per institutional standard on Days 1 and 8
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following gemcitabine
- Regimen 4 (non-squamous histology only):
  - Pemetrexed 500 mg/m^2 IV over 10 minutes or per institutional standard on Day 1
  - Cisplatin 75 mg/m^2 IV over 120 minutes or per institutional standard on Day 1, immediately following pemetrexed
- Regimen 5:
  - Paclitaxel 175 or 200 mg/m^2 IV over 180 minutes or per institutional standard on Day 1
  - Carboplatin AUC5 or 6 IV over 30 minutes or per institutional standard on Day 1, immediately following paclitaxel

The study design schematic is presented in Figure 2.1-1.

Figure 2.1-1: Study Design Schematic



A Data Monitoring Committee (DMC) will be established to provide oversight of safety and efficacy and overall risk/benefit monitoring of the study.

Note that in this document the words "participant" and "subject" are used intercheangeably.

2.2 Treatment Assignment

CA209816 is an open-label, randomized trial. Participants with Stage IB (≥ 4 cm), II and IIIA (N2) considered resectable will be eligible to participate. After the participant's initial eligibility is established and informed consent has been obtained, the participant must be enrolled into the study by calling the IRT to obtain a participant number. Every participant that signs the informed consent form must be assigned a participant number in IRT.

Once enrolled in IRT, enrolled participants who have met all eligibility criteria will be ready to be randomized through IRT to treatment Arm A, Arm B or Arm C.

In the original study design (before revised protocol 02) subjects were randomized to in a 1:1 ratio to arms A or B.

Following revised protocol 02, subjects were randomized to in a 1:1:1 ratio to arms A, B or C.

As of revised protocol 03, subjects were randomized to in a 1:1 ratio to arms B or C.

- Arm A: nivolumab 3 mg/kg plus ipilimumab 1 mg/kg.
- Arm B: platinum doublet chemotherapy.
- Arm C: nivolumab 360mg flat dose plus platinum doublet chemotherapy

The randomization uses permuted blocks stratified by the following factors:

- PD-L1 expression (≥1% or <1%/not evaluable/indeterminate)
- Disease stage (IB/II vs IIIA)
- Gender

2.3 Blinding and Unblinding

This is an open-label study; blinding procedures between participants and investigators are not applicable. The specific treatment to be taken by a participant will be assigned using an IRT. No aggregate summary data by treatment group are disclosed to the study team at any time of the study conduct until achievement of primary endpoint significance (pCR).

As described in the DMC charter, at the time of the first EFS interim analysis, in case EFS is not significant, the closed DMC report (unblinded) will still be shared with a BMS executive restricted team (3 persons named in DMC charter) to possibly engage in conversation with health authorities based on a trend in EFS (even if not significant), in the context of a statistically significant result in the pCR primary endpoint.

After discussion with health authorities, the BMS Executive restricted team may decide to share the DMC closed report with limited additional BMS members in order to prepare a broader health authority interaction. These additional BMS members will be outside of the CA209816 study team in addition to a small number of oncology leaders who would remain firewalled to the study in case the decision following health authorities interaction would be not to proceed to application. The members of this health authority interaction preparation team will be documented in a BMS internal document prior to the DMC meeting.

After formal discussion with FDA, should BMS proceed with an application for registration, the BMS study team will be unblinded to the study results and data.

Treatment assignments will be released to the bioanalytical laboratory in order to minimize unnecessary analysis of samples.

The blinded independent pathology review (BIPR) and blinded independent central review (BICR) will be blinded to treatment arms.

2.4 Protocol Amendments

Table 2.4-1: Protocol Amendments

Document/Date of Issue	Summary of Change
Revised Protocol 07 Under finalization	To account for potential slowdown in event-free survival (EFS) events accrual in long-term follow-up the revised protocol was updated to: • Revise EFS modeling assumptions to a piecewise exponential with lower events rate in the longer term. • Include one additional EFS interim analysis at 90% information fraction [redacted]
Revised Protocol 06	• Clarified EFS definition

Table 2.4-1: Protocol Amendments

Document/Date of Issue	Summary of Change
14-Jul-2020	• Removed the 1st interim analysis of EFS and updated alpha spending of interim and final analyses of EFS • Clarified that actual timing of analyses may differ from projected timin • Removed text about descriptive EFS analysis [redacted]
Revised Protocol 05 18-Sep-2019	• Modified pCR analysis population and projected timelines. – Rationale: based on FDA feedback indicating that the subset of 260 patients might not yield adequate numbers of patients with a pCR upon which to base reasonable assumptions of clinical meaningfulness. • Updated surgical approach endpoint • Updated the censoring rule of TTDM • No optional biopsy at disease progression collected in China. • Updated Management Algorithms to include myocarditis
Revised Protocol 04 25-Jun-2019	[redacted] • Added the concomitant administration of substances that are also tubularly secreted (eg, probenecid) could potentially result in delayed clearance of pemetrexed. • Added hypothesis testing for overall survival • Clarified the pCR analysis population • Added exploratory endpoint of Event Free Survival on next line of therapy • Added instructions for BICR • Updated Appendix 8 Adverse Events and Serious Adverse Events: Definitions and Procedures for Recording, Evaluating, Follow Up and Reporting Adverse Events
Revised Protocol 03 21-Sep-2018	• Enrollment in Arm A (nivolumab plus ipilimumab) was stopped. Based on external data from a neo-adjuvant phase II trial (NADIM[4]) which suggested anti-PD-1 + chemotherapy to be more promising, clinical development of nivolumab + chemotherapy was prioritized. • Randomization of participants (1:1 ratio) into Arm C and B will continue to for a total of 350 participants

Table 2.4-1: Protocol Amendments

Document/Date of Issue	Summary of Change
	• Definition of event free survival is clarified. • Participants with large-cell neuroendocrine carcinoma tumor histology are excluded • Additional platinum doublet chemotherapy regimen (paclitaxel/carboplatin) was added • Dose modification for docetaxel was updated • Time to death or distant metastases (TTDM) was added to secondary endpoints. • Tumor assessments for participants who do not proceed to definitive surgery was clarified • Endpoints and statistical analyses adapted consequently to Arm A discontinuation. • Rationale, background information, and trial schematic were updated • Pulmonary function parameters were clarified • Time relationship between adjuvant radiotherapy and tumor imaging assessments was clarified • Time window of Cycle 1 Day 1 end-of-infusion PK sampling for Arm A and C was clarified
Revised Protocol 02 06-Jul-2017	• Nivolumab plus platinum-doublet chemotherapy arm (Arm C) was added. • The sample size was increased to 642 participants consequently to the addition of Arm C and change of primary endpoint. • The primary objective was changed to multiple primary objectives of event free survival and pathological complete response; major pathological response was changed to the secondary objective. • Additional rationale and background information was provided. • Pre-screening tissue requirement was increased from minimum of 10 slides to 15 slides. • Contrast requirements for brain MRI scans were updated. • Time window for pulmonary function test window was expanded from within 28 days of randomization to within 6 weeks of randomization. • Synopsis was updated with exploratory objectives, endpoints and schema. • Language in treatment administered was deleted and reference to Investigator Brochure and Pharmacy Manual was included. [redacted]
Revised Protocol 01	• Incorporates Amendment 02 and Administrative Letters 01 and 02

Table 2.4-1: Protocol Amendments

Document/Date of Issue	Summary of Change
03-Mar-2017	
Amendment 02 03-Mar-2017	• To adjust the dosing details of the chemotherapy regimens to include the dose approved by the local prescribing information and the standard of care infusion time for each country included in this study. • To expand and to split the broad biomarker objective into 3 more detailed objectives. • Clarify lymph node samples at screening and at definitive surgery. • Clarify requirements for PET/CT scans and broadening the window of scans prior to surgery. • Clarify tissue sample process for calculation of the primary endpoint. • Adjust Hepatitis B Virus criteria. • Added live vaccines and strong CYP3A4 inhibitors to the Prohibited Treatments. • added caution for concomitant administration of NSAIDs with pemetrexed • added unacceptable methods of contraception to Appendix 6.
Administrative Letter 02 30-Nov-2017	• Clarify the correct version of the TNM Staging System. • Clarify that a minimum of 228 PD-L1+ participants will be randomized. • Clarify that physical exams, vital signs, and physical measurements should be collected prior to each dose of neoadjuvant and adjuvant therapy. • Clarify that the first post-operative tumor assessment should be performed 12 weeks (± 7 days) after definitive surgery. • Remove the phrase "non-protocol regimen" in regards to a noncisplatin • Regimen as the protocol includes a non-cisplatin regimen option. • Clarify that weight-based dosing should be rounded up to the nearest milligram or per institutional standards. • Clarify that the EQ-5D-3L should be collected prior to Day 1 only in cycles that have multiple dosing days in each cycle.
Administrative Letter 01 31-Oct-2017	• To correct the IND number
Original Protocol	• Not Applicable

Table 2.4-1: Protocol Amendments

Document/Date of Issue	Summary of Change
30-Sep-2016	

2.5 Data Monitoring and Other External Committees

A Data Monitoring Committee (DMC) is established to provide oversight of safety and efficacy considerations in protocol CA209816. Additionally, the DMC will provide advice to the sponsor regarding actions the committee deems necessary for the continuing protection of participants enrolled in the study. The DMC will be charged with assessing such actions in light of an acceptable benefit/risk profile for nivolumab in combination with ipilimumab or chemotherapy. The DMC will act in an advisory capacity to BMS and will monitor participant safety and evaluate the available efficacy data for the study. The oncology therapeutic area of BMS has primary responsibility for design and conduct of the study.

Independent pathology (BIPR) and radiology review (BICR) will be established for central review and confirmation of efficacy endpoints.

3 OBJECTIVES

3.1 Primary

- To compare the event-free survival (EFS) by BICR in participants receiving nivolumab plus platinum doublet chemotherapy vs participants receiving platinum doublet chemotherapy in operable stage IB (≥ 4 cm), II, or resectable IIIA (N2) NSCLC
- To compare the pathologic complete response (pCR) rate in participants receiving nivolumab plus platinum doublet chemotherapy vs participants receiving platinum doublet chemotherapy in operable stage IB (≥ 4 cm), II, or resectable IIIA (N2) NSCLC

3.2 Secondary

- To assess the major pathologic response (MPR) rate by BIPR of participants receiving nivolumab plus platinum doublet chemotherapy vs participants receiving platinum doublet chemotherapy in operable stage IB (≥ 4 cm), II, or resectable IIIA (N2) NSCLC
- To assess the OS of participants receiving nivolumab plus platinum doublet chemotherapy vs participants receiving platinum doublet chemotherapy in operable stage IB (≥ 4 cm), II, or resectable IIIA (N2) NSCLC
- To assess the time to death or distant metastases (TTDM) of participants receiving nivolumab plus platinum doublet chemotherapy vs participants receiving platinum doublet chemotherapy in operable stage IB (≥ 4 cm), II, or resectable IIIA (N2) NSCLC

3.3 Exploratory Objectives

- To assess clinical response rate (cRR) by BICR of participants receiving nivolumab plus platinum doublet chemotherapy vs participants receiving platinum doublet chemotherapy in operable stage IB (≥ 4 cm), II, or resectable IIIA (N2) NSCLC

- To assess the pCR rate, MPR rate , cRR, EFS, TTDM and OS in early-stage NSCLC participants treated with nivolumab plus platinum doublet chemotherapy compared to those treated with platinum doublet chemotherapy by PDL1 status (PD-L1 ≥ 1% , PD L1< 1% /not evaluable/ indeterminate)
- To assess the feasibility of surgery and rate of peri- and post-operative complications (within 90 days of surgery) in participants receiving nivolumab plus platinum doublet chemotherapy compared to participants receiving platinum doublet
- To assess the safety and tolerability of nivolumab plus platinum doublet chemotherapy compared to platinum doublet chemotherapy in early stage NSCLC
- To describe the pCR rate, MPR rate, cRR, EFS, OS, TTDM, feasibility of surgery, rate of peri- and post-operative complications (within 90 days of surgery), safety and tolerability in early-stage NSCLC participants treated with nivolumab plus ipilimumab and by PDL1 status (PD-L1 ≥ 1% , PD L1< 1%/not evaluable/indeterminate)
- To assess pharmacokinetics of the nivolumab plus ipilimumab or nivolumab plus platinum doublet chemotherapy in participants with early stage NSCLC
- To assess the participant's overall health status and health utility using the 3-level version of the EQ-5D-3L visual analog scale (VAS) and utility index, respectively
- To evaluate tumor mutational burden as a potential predictive biomarker of efficacy (such as EFS and OS) of nivolumab plus platinum doublet chemotherapy and of platinum-doublet chemotherapy, using data generated from tumor and blood (germ-line control) specimens.

[redacted]

- To explore potential predictive biomarkers of nivolumab plus platinum doublet chemotherapy efficacy (such as EFS and OS) in peripheral blood and tumor specimens [redacted]

4 ENDPOINTS

4.1 Primary Endpoint(s)

The primary objectives in the study will be evaluated by the multiple primary endpoints of EFS and pCR.

4.1.1 Event-Free Survival

Two definitions are used for analysis of EFS. The primary definition accounts for subsequent therapy by censoring at the last evaluable tumor assessment on or prior to the date of subsequent therapy (outside of the protocol specified adjuvant therapy). The secondary definition does not incorporate censoring due to subsequent therapy.

EFS rate at time *T* is defined as the probability that a subject has not progressed/recurred and is alive at time *T* following randomization. EFS rates at fixed time points (e.g. 12 months, depending on the minimum follow-up) are defined as the probability that a subject has not progressed and is alive at time *T* following randomization.

4.1.1.1 Primary Definition of Event-Free Survival

Event free survival is defined as the length of time from randomization to any of the following events: any progression of disease precluding surgery, progression or recurrence disease after surgery (based on BICR assessment per RECIST 1.1), or death due to any cause.

- A pre-surgical progression (even if reaching the RECIST 1.1 criteria) which does not preclude surgery is not considered as an event.
- A progression not reaching the RECIST 1.1 criteria (e.g. clinical progression) but which still precludes surgery (i.e. reason for no surgery is disease progression) is considered as an event (event at the investigator reported earliest clinical or radiographic progression date, or at the date of randomization if no progression date reported).
- For participants with surgery, any new lesions identified by BICR on the first post-surgical baseline imaging compared with the pre-surgical scans will be identified as new lesion and will be counted as an event. For those without new lesion on the first post-surgical scan, the first tumor assessment post surgery will be used as re-baseline and recurrence/progression per RECIST 1.1 by BICR will be evaluated based on that re-baseline.
- Participants who do not undergo surgery for reason other than progression will be considered to have an event at RECIST 1.1 progression (based on BICR) or death.
- Participants who died without a reported progression/disease recurrence will be considered to have experienced an event on the date of their death.

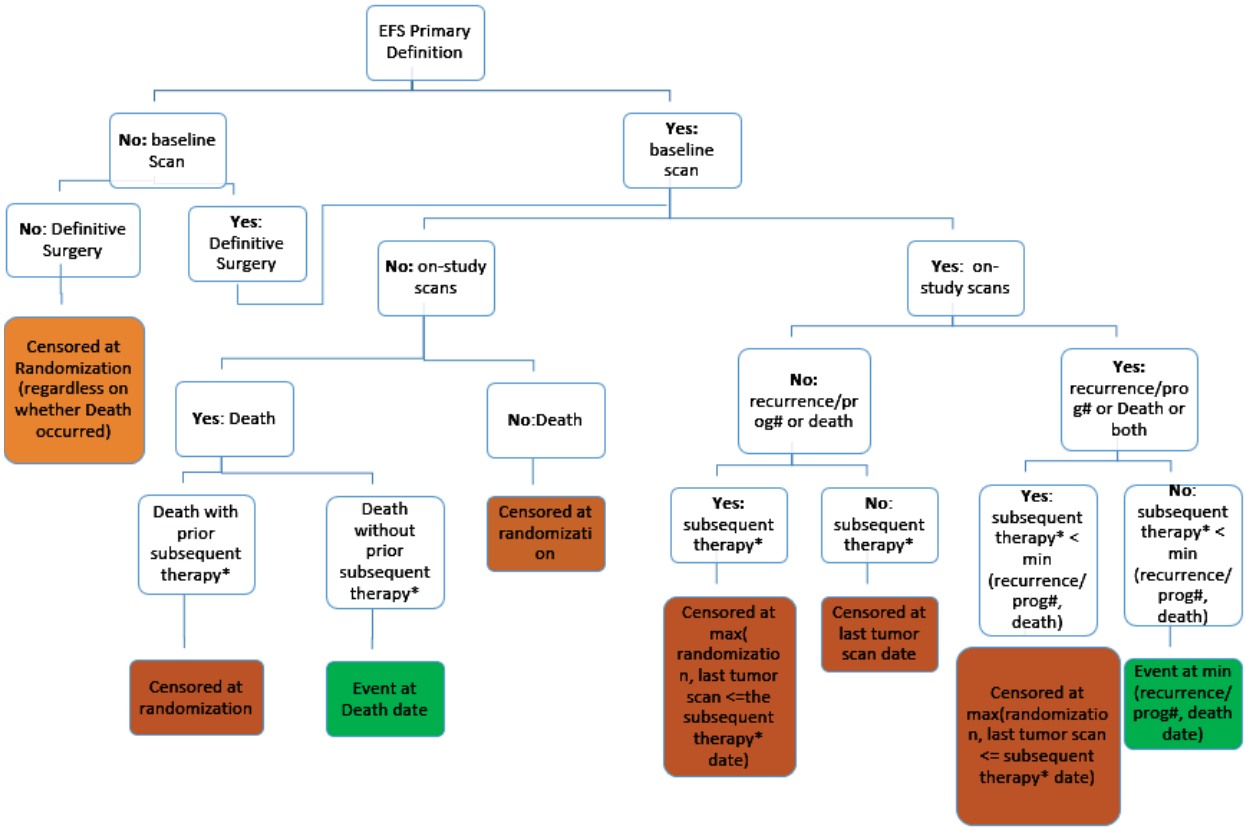
The following censoring rules will be applied for the primary definition of EFS:

- Participants who did not report progression/recurrence of disease or die will be censored on the date of their last evaluable tumor assessment.
- Participants who did not have any on study tumor assessments and did not die will be censored on the date they were randomized.
- Subjects who receive subsequent anti-cancer therapy, outside of the protocol-specified adjuvant therapy, prior to documented progression/recurrence/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.
- Subjects who did not have a documented progression/recurrence/death and received subsequent anti-cancer therapy outside of the protocol-specified adjuvant therapy 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.

- Participants without baseline scan and without surgery will be censored on the date of randomization (regardless of death).
- Censoring rules for the primary definition of EFS (EFS truncated at subsequent therapy) are presented as follows and depicted in Figure 4.1.1.1-1.
- It is to be noted that in case of new primary cancer, if such lesions are present on tumor assessment at the BICR, they will be considered as new lesions, since the BICR does not have access to biopsy results.

Figure 4.1.1.1-1: EFS Primary Definition



*Subsequent Therapy excluding per protocol adjuvant therapy

Progression precluding surgery, RECIST 1.1 recurrence or progression post surgery (for participants with surgery), RECIST 1.1 progression (for participants without surgery)

4.1.1.2 Secondary Definition of Event-Free Survival

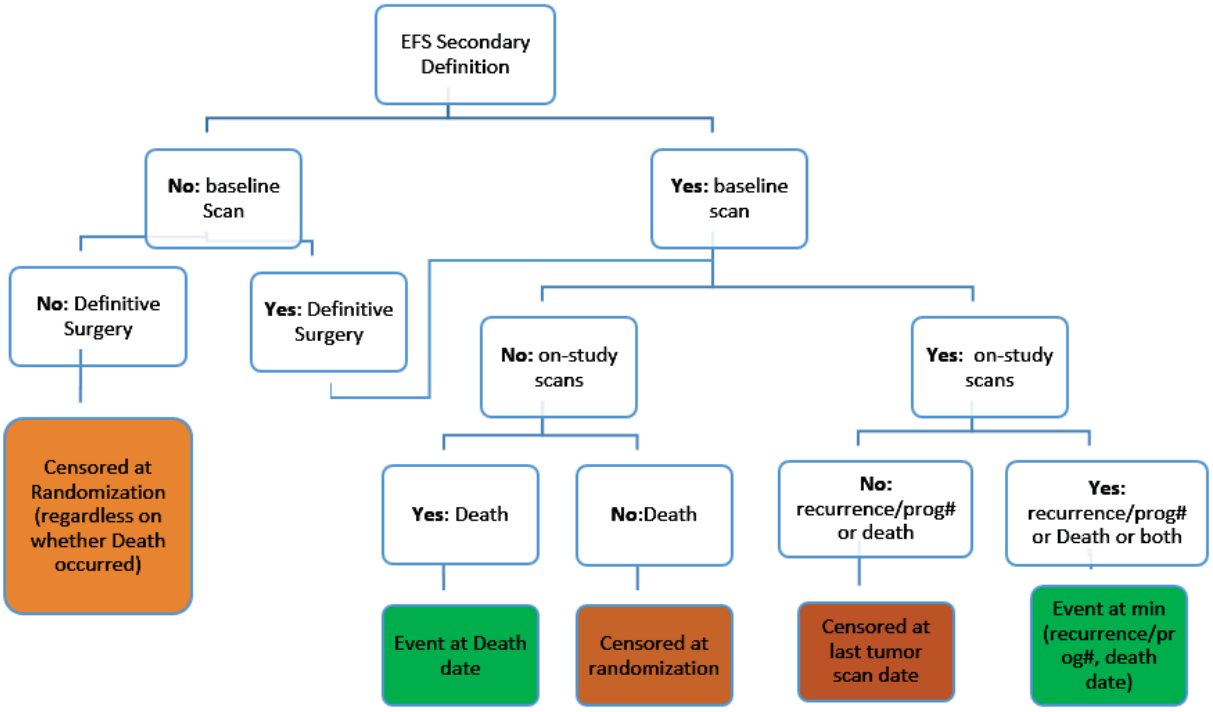
The secondary definition of EFS (ITT definition) is defined as the length of time from randomization to any of the following events: any progression of disease precluding surgery, progression or recurrence disease after surgery (based on BICR assessment per RECIST 1.1), or death due to any cause.

- Progression/recurrence will be based on BICR assessment per RECIST 1.1.
- A progression (even if reaching the RECIST 1.1 criteria) which does not preclude surgery is not considered as an event.
- A progression not reaching the RECIST 1.1 criteria but which still precludes surgery (i.e. reason for no surgery is disease progression) is considered as an event (event at the investigator reported earliest clinical or radiographic progression date, or at the date of randomization if no progression date reported).
- For participants with surgery, the first tumor assessment post surgery will be used as re-baseline and recurrence/progression per RECIST 1.1 will be evaluated based on that re-baseline. Any new lesions on the post-surgical baseline imaging compared with the pre-surgical scans will be identified as new lesion and will be counted as an event.
- Participants who do not undergo surgery for reason other than progression will be considered to have an event at RECIST 1.1 progression or death.
- Participants who died without a reported progression/disease recurrence will be considered to have experienced an event on the date of their death.

The following censoring rules will be applied for the secondary definition of EFS:

- Participants who did not report progression/recurrence of disease or die will be censored on the date of their last evaluable tumor assessment.
- Participants who did not have any on study tumor assessments and did not die will be censored on the date they were randomized.
- Participants without baseline scan and without surgery will be censored on the date of randomization (regardless of death).
- Censoring rules for the secondary definition of EFS (ITT definition) are presented as follows and depicted in Figure 4.1.1.2-1

Figure 4.1.1.2-1: EFS Secondary Definition



Progression precluding surgery, RECIST 1.1 recurrence or progression post surgery (for participants with surgery), RECIST 1.1 progression (for participants without surgery)

4.1.2 Pathologic Complete Response Rate

Pathological complete response (pCR) rate is defined as number of randomized participants with absence of residual tumor in lung and lymph nodes at surgery as evaluated by blinded independent pathological review (BIPR), divided by the number of randomized participants for each treatment group. Randomized subjects who are no longer eligible for surgery, or who are on alternative anti-cancer therapy before surgery, or who discontinue the study (e.g. withdraw consent) before surgery, or who otherwise do not have an evaluable BIPR result available are all counted as non-responders.

4.2 Secondary Endpoint(s)

4.2.1 Overall Survival

Overall survival (OS) is defined as the time between the date of randomization and the date of death due to any cause. For a subject without documentation of death, OS will be censored on the last date the subject was known to be alive.

4.2.2 Major Pathological Response Rate

Major pathological response (MPR) rate, defined as number of randomized participants with ≤ 10% residual tumor in lung and lymph nodes at surgery as evaluated by BIPR, divided by the number of randomized participants for each treatment group. Viable tumors in situ carcinoma should not be included in MPR calculation. Randomized subjects who are no longer eligible for

surgery, or who are on alternative anti-cancer therapy, or who discontinue the study (e.g. withdraw consent) before surgery, or who otherwise do not have an evaluable BIPR result available are all counted as non-responders.

4.2.3 Time to Death or Distant Metastases

Time to Death or Distant Metastates (TTDM) is defined as the time between the date of randomization and the first date of distant metastasis or the date of death in the absence of distant metastasis. Distant metastasis is defined as any new lesion that is outside of the thorax using BICR according to RECIST 1.1. It will be derived based on the location of lesions outside the thorax. Participants who died without reported distant metastasis will be considered to have experienced an event on the date of their death.

The following censoring rules will be applied TTDM:

- Participants who have not developed distant metastasis nor died will be censored on the date of their last evaluable tumor assessment.
- Participants who did not have any on study tumor assessments and did not die will be censored on the date they were randomized.

4.3 Exploratory Endpoint(s)

4.3.1 Clinical Response Rate by BICR

Clinical response rate (cRR) is defined as proportion of randomized participants whose overall radiological response prior to definitive surgery (or best overall radiological response (BOR) at the first protocol planned tumor assessment if a subject has no surgery) is either a complete response (CR) or partial response (PR) per RECIST 1.1 criteria by BICR. Participants who received alternative anti-cancer therapy before the pre-surgery tumor assessment will be counted as non-responders.

4.3.2 Event Free Survival on Next Line of Therapy (EFS2)

EFS on next line therapy (EFS2) is defined as the 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. Subjects 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 EFS2:

- Subjects who did not receive subsequent next line systemic anti-cancer therapy:
  - Subjects who died, the death date is the event date;
  - Else the subject's EFS2 is censored at the last known alive date.
- Subjects who received subsequent next line anti-cancer therapy:
  - Subjects who had a disease recurrence/progression after the start of subsequent anti-cancer therapy, this disease progression date is the event date;
  - Else if a subject died or start of second next line therapy, the date of min (death, start date of second next line therapy) is the event date;

– Else the subject's EFS2 is censored at the last known alive date.

Subsequent next line of therapy will include subsequent systemic regimen given in one of the following settings Unresectable, Locally Advanced or lines of therapy in metastatic setting.

4.3.3 Surgery Related Endpoints

The endpoints related to surgery include proportion of subjects with delayed (including duration of delay) or canceled surgery, duration of surgery, length of hospital stay, surgical approach, including completeness of surgery (R0/R1/R2 resection), incidence of AE/SAE associated with surgery up to 90 days after surgery.

4.3.4 Safety and Tolerability

The assessment of safety will be based on the incidence of adverse events (AEs), serious adverse events (SAEs), adverse events leading to discontinuation, adverse events leading to dose modification, select adverse events (select AEs) for EU/ROW Submissions, immune-mediated AEs (IMAEs) for US Submission, other events of special interest (OEOSI), and deaths. The use of immune modulating concomitant medication will be also summarized. In addition clinical laboratory tests will be analyzed.

4.3.5 Pharmacokinetics

Pharmacokinetics will be measured by the serum concentration of nivolumab and ipilimumab. Samples will be collected to characterize pharmacokinetics of nivolumab and ipilimumab and to explore exposure-safety and exposure-efficacy relationships. The population pharmacokinetics analysis will be presented separately from the main clinical study report.

4.3.6 Biomarkers

Biomarkers potentially associated with clinical endpoints will be measured by analyzing tumor and blood samples.

Biomarker endpoints include, but not limited to, tumor mutational burden (TMB) using data generated from tumor specimens., tumor inflammatory gene expression signatures using data generated from tumor specimen [redacted] Results for biomarkers analyses (other than PD-L1, TMB [redacted]) will be summarized outside of CSR.

4.3.6.1 Tumor Mutational Burden

TMB is measured in CA209816 using the [redacted] assay. [redacted] is a next-generation sequencing (NGS) assay targeting the full coding regions of 523 genes implicated in the pathogenesis of solid tumors. Using enrichment-based library preparation techniques for use with formalin-fixed, paraffin-embedded (FFPE) samples, [redacted] can analyze DNA and RNA from the same sample, detecting single nucleotide variants (SNVs), insertions and deletions (indels), amplifications, splice variants, and fusions, in a single sequencing run. TMB, is derived by summing the total of all synonymous and non-synonymous detected small DNA

variants (SNVs and indels) across the entire coding region (~1.3Mb are in coding regions) with sophisticated variant calling and germline filtering algorithms for enhanced accuracy. The resulting number is communicated as mutations per Mb unit (mut/Mb). The cutoff used for analysis will be ≥ 12.3 mut/Mb, <12.3 mut/Mb[5].

4.3.6.2 PD-L1 Protein Expression

PD-L1 expression is defined as the percent of tumor cells membrane staining in a minimum of 100 evaluable tumor cells per validated Dako PD-L1 immunohistochemistry (IHC) assay. This is referred to as quantifiable PD-L1 expression. If the PD-L1 staining could not be quantified, it is further classified as:

1) Indeterminate: Tumor cell membrane staining hampered for reasons attributed to the biology of the tumor tissue sample and not because of improper sample preparation or handling.
2) Not evaluable: Tumor tissue sample was not optimally collected or prepared and PD-L1 expression is neither quantifiable nor indeterminate. Not evaluable can be determined from H&E process before the tumor biopsy specimen is sent for PD-L1 evaluation or from the H&E process during PD-L1 evaluation.

Subjects with missing PD-L1 expression are subjects with no tumor tissue sample available for evaluation.

PD-L1 expression will be collected in the IRT as well as in the clinical database. Statistical analysis using PD-L1 expression will be solely based on PD-L1 expression data from clinical database. Stratified analyses will use stratification from IRT, unless otherwise specified.

4.3.7 Outcomes Research

4.3.7.1 EQ-5D-3L

Subjects' reports of general health status will be assessed using the EuroQoL Group's EQ-5D-3L. EQ-5D-3L essentially has 2 components: the descriptive system and the visual analogue scale (VAS).

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, described by the EQ-5D-3L. Altogether, the instrument describes $3^5 = 243$ health states. 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. For this study, EQ-5D-3L utility index values will be computed using a scoring algorithm based on the United Kingdom Time-Trade-Off (UK TTO) value set[6]

In addition, the EQ-5D-3L includes a VAS, which allows respondents to rate their own current health on a 101-point scale ranging from 0=”worst imaginable” health to 100=“best imaginable” health state [7].

All questionnaires completed at baseline and on-study will be assigned to a time-point according to the windowing criteria in Table 4.3.7.1-1 and included in the analysis. In case a subject has two on-study assessments within the same window, the assessment closest to the time-point will be used and, in the case of two assessments at a similar distance to the time-point, the latest one will be chosen. In the event where the subject has no assessment at all in a specific window, the observation will be treated as missing for that time-point.

Table 4.3.7.1-1: Time Windows for EQ-5D-3L Assessments

Time Point	Nominal Day	Time Window
Baseline	D1	Prior to first dose on Day 1
Week 3 (Arm A)	D15	Day 2 thru day 22 inclusive
Week 4 (Arms B and C)	D22	Day 2 thru day 32 inclusive
Week 5 (Arm A)	D29	Day 23 thru day 36 inclusive
Week 7 (Arms B and C)	D43	Day 33 thru day 53 inclusive
Post-neoadjuvant visit 1	Last neodajuvant dose + 30 days	Assessment post last neoadjuvant dose and within 65 days of last neo dose
Post-neoadjuvant visit 2	Last neodajuvant dose + 100 days	Assessment post 65 days of last neoadjuvant dose and within 145 days of last neo dose
Subjects without adjuvant:		
Survival Follow-up 1	Last neodajuvant dose + 190 days	Assessment post 145 days of last neoadjuvant dose and within 235 days of last neo dose
Survival Follow-up 2	Last neodajuvant dose + 280 days	Assessment post 235 days of last neoadjuvant dose and within 325 days of last neo dose
Survival Follow-up 3	Last neodajuvant dose + 370 days	Assessment post 325 days of last neoadjuvant dose and within 415 days of last neo dose
Survival Follow-up 4	Last neodajuvant dose + 460 days	Assessment post 415 days of last neoadjuvant dose and within 550 days of last neo dose
Survival Follow-up 5	Last neodajuvant dose + 640 days	Assessment post 550 days of last neoadjuvant dose and within 730 days of last neo dose
Survival Follow-up 6	Last neodajuvant dose + 820 days	Assessment post 730 days of last neoadjuvant dose and within 910 days of last neo dose
Then Survival Follow-up i	Last neodajuvant dose + 460 + (i-4)*180 days	Assessment post (nominal day - 90) days of last neoadjuvant dose and within (nominal day + 90) days of last neo dose
Subjects with systemic adjuvant:		
Adjuvant Cycle 1	-	Assessment reported in the F01 (adjuvant cycle 1) visit

Table 4.3.7.1-1: Time Windows for EQ-5D-3L Assessments

Time Point	Nominal Day	Time Window
Adjuvant Cycle 2	-	Assessment reported in the F02 (adjuvant cycle 2) visit
Adjuvant Cycle 3	-	Assessment reported in the F03 (adjuvant cycle 3) visit
Adjuvant Cycle 4	-	Assessment reported in the F04 (adjuvant cycle 4) visit
Survival Follow-up 1	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 90 days	Assessment post 45 days of Max (Last adjuvant dose or Post-neoadjuvant visit) and within 135 days of Max (Last adjuvant dose or Post-neoadjuvant visit)
Survival Follow-up 2	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 180 days	Assessment post 135 days of Max (Last adjuvant dose or Post-neoadjuvant visit) and within 225 days of Max (Last adjuvant dose or Post-neoadjuvant visit)
Survival Follow-up 3	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 270 days	Assessment post 225 days of Max (Last adjuvant dose or Post-neoadjuvant visit) and within 315 days of Max (Last adjuvant dose or Post-neoadjuvant visit)
Survival Follow-up 4	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 360 days	Assessment post 315 days of Max (Last adjuvant dose or Post-neoadjuvant visit) and within 450 days of Max (Last adjuvant dose or Post-neoadjuvant visit)
Survival Follow-up 5	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 540 days	Assessment post 450 days of Max (Last adjuvant dose or Post-neoadjuvant visit) and within 630 days of Max (Last adjuvant dose or Post-neoadjuvant visit)
Survival Follow-up 6	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 720 days	Assessment post 630 days of Max (Last adjuvant dose or Post-neoadjuvant visit) and within 810 days of Max (Last adjuvant dose or Post-neoadjuvant visit)
Then Survival Follow-up i	Max (Last adjuvant systemic dose or Post-neoadjuvant visit) + 360 + (i-4)*180 days	Assessment post (nominal day - 90) days of last neo dose and within (nominal day + 90) days of last neo dose

5 SAMPLE SIZE AND POWER

The original study design (before Revised protocol 02) had two arms, with participants randomized in a 1:1 ratio to either neoadjuvant nivolumab plus ipilimumab or platinum doublet chemotherapy arm. Revised protocol 02 added a new, neoadjuvant nivolumab plus platinum doublet chemotherapy arm. When the third arm opens and as each site receives IRB/EC approval of revised

protocol 02, the IRT will switch to a 1:1:1 randomization at the respective site. Starting from that point on, the site will only enroll under revised protocol 02.

Revised protocol 03 withholds randomization into the arm of neoadjuvant nivolumab plus ipilimumab but continues randomizing eligible participants into either neoadjuvant nivolumab plus platinum doublet chemotherapy arm or platinum doublet chemotherapy arm in a 1:1 ratio. Approximately 350 participants (175 participants per arm) will be randomized between 2 arms neoadjuvant nivolumab plus platinum doublet chemotherapy or platinum doublet chemotherapy from 1:1:1 randomization in revised protocol 02 and 1:1 randomization in revised protocol 03. Participants already randomized in the original 2-arm part (neoadjuvant nivolumab plus ipilimumab vs neoadjuvant chemotherapy) and in the arm of neoadjuvant nivolumab plus ipilimumab in 3-arm part defined by revised protocol 02 will remain in trial and continue scheduled trial procedures. It is expected to have around 70 participants randomized in the original 2-arm part and approximately other 75 participants randomized in the arm of neoadjuvant nivolumab plus ipilimumab in the 3-arm part. It is estimated that there will be a total of approximately 500 participants on the study.

Starting from 1:1:1 randomization, approximately 350 participants will be randomized to the 2 arms neoadjuvant nivolumab plus platinum doublet chemotherapy or platinum doublet chemotherapy in a 1:1 ratio (concurrently randomized).

The sample size of the study is calculated based on the primary endpoint of EFS and accounts for the multiple primary endpoints comparisons: pCR (per BIPR) and EFS (per BICR) with an initial alpha allocation of 0.01 and 0.04 respectively. Formal analyses of pCR and EFS may be conducted at different timepoints. The fallback method will be used, ie, if the pCR comparison between Arm C and Arm B is statistically significant, then 0.01 alpha allocated to pCR will be passed to the EFS comparison for Arm C vs Arm B and the EFS comparison will be conducted at the alpha = 0.05 level. If the pCR comparison between Arm C and Arm B is not statistically significant, then the EFS comparison for Arm C vs Arm B will be conducted at the alpha = 0.04 level. [redacted]

5.1 Pathologic Complete Response (pCR)

The primary analysis of pCR will be performed after the 350 randomized participants in neoadjuvant nivolumab plus platinum doublet chemotherapy and platinum doublet chemotherapy (from start of 1:1:1 randomization) have an opportunity for surgery.

Assuming an accrual rate of 10 participants (all comers) a month between Arms B and C during 1:1:1 randomization (about 10 months), and 15 participants per month during 1:1 randomization, it is anticipated that the 350 participants will be randomized in approximately 27 months. The pCR endpoint is expected to be analyzed after about 30 months from start of 1:1:1 randomization.

Assuming pCR rate of 10% on Arm B chemotherapy and 30% on Arm C nivolumab plus chemotherapy, respectively, the 350 participants will provide more than 90% power to detect an odds ratio of 3.857 with a 2-sided type I error of 1%.

It is estimated that there will be about 110 subjects randomized to Arm A neoadjuvant nivolumab plus ipilimumab before revised protocol 03 is implemented. Assuming true pCR rate is 15% on this arm, there is 95% probability that the lower bound of 95% exact confidence interval of pCR is above 5%.

5.2 Event Free Survival (EFS)

For the formal comparison of EFS as assessed by BICR for nivolumab plus platinum doublet chemotherapy (Arm C) vs platinum doublet chemotherapy (Arm B), only participants randomized from 1:1:1 randomization in revised protocol 02 and 1:1 randomization in revised protocol 03 will be counted (participants concurrently randomized in arms B and C).

Considering this SAP revision (v3) occurs after the readout of the pCR primary endpoint which was statistically significant, the power details are provided below using alpha=0.05 (0.01 alpha from the pCR endpoint fallback to the EFS comparison. In addition, it reflects the number of subjects that were actually concurrently randomized in Arms B and C: 358 subjects.

A total of 185 events ensure that an overall 2-sided 5% significance level sequential test procedure with two interim analyses after 148 events (80% of events required for final analysis) and 167 events (90% of events required for final analysis) in 358 randomized participants will have 82% power assuming an HR of 0.65 between the 2 Arms. Considering a piecewise exponential distribution with control hazard rates of 0.028 before 20 months, 0.017 between 20 months and 40 months, 0.014 between 40 and 60 months and 0.008 after 60 months, and a dropout rate of approximately 20%, it is anticipated that the EFS analyses will take place at about 48, 58, and 73 months from start of 1:1:1 randomization. The trigger of the first interim analysis is event driven. The second interim analysis will take place when 167 events are observed [redacted]. The final analysis will take place when approximately 185 events are observed [redacted]. The stopping boundaries at the interim and final EFS analyses will be derived based on the exact number of events using Lan-DeMets alpha spending function with O'Brien-Fleming boundaries. If the interim analyses of EFS are performed at exactly 148 and 167 events, the nominal significance level for EFS superiority will be 0.024 and 0.030, respectively. The nominal significance level for the final look of EFS after 185 events would then be 0.038.

[redacted]

Table 5.3-1 summarizes the key parameters of the sample size justification in the concurrently randomized participants from Arms B and C.

5.3 Power Considerations for Overall Survival

The secondary endpoint Overall survival will be tested hierarchically after EFS with the same overall alpha as for the EFS comparison (two-sided 4% if the pCR comparison is not significant or 5% if the pCR comparison is significant).

For the formal comparison of OS for nivolumab plus platinum doublet chemotherapy (Arm C) vs platinum doublet chemotherapy (Arm B), only participants concurrently randomized from 1:1:1 randomization in revised protocol 02 and 1:1 randomization in revised protocol 03 will be included. [redacted]

Considering this SAP revision (v3) occurs after the readout of the pCR primary endpoint which was statistically significant, the power details are provided below using alpha=0.05 (0.01 alpha from the pCR endpoint fallback to the EFS comparison, then hierarchically on the OS comparison). In addition, it reflects the number of subjects that were actually concurrently randomized in Arms B and C: 358 subjects.

[redacted]

[redacted] The stopping boundaries at the interim and final OS analyses will be derived based on the exact number of events using Lan-DeMets alpha spending function with O'Brien-Fleming boundaries. This spending function is specific to OS and accounts for potential interim OS analyses even if they did not actually take place because of EFS non-significance[8].

[redacted]

Table 5.3-1 summarizes the key parameters of the power calculation for EFS and OS in the concurrently randomized participants from Arms B and C.

Table 5.3-1: Power Calculation for EFS and OS

	EFS **Arm C vs Arm B**	**OS** **Arm C vs Arm B**
Accrual	Actual accural 25 months	Actual accural 25 months
Power	82%	[redacted]
Two-sided alpha	0.05	0.05
Hypothesized Median Control vs exp (months)	28 vs 52* Piecewise exponential model	[redacted]
Hypothesized Hazard ratio	0.65	0.65
Sample size for concurrent comparison	358	358
First interim analysis for EFS (EFS IA1) and OS (OS IA1)	148 events Alpha boundary: 0.024	Triggered by EFS IA1 [redacted]
Second interim analysis for EFS (EFS IA2) and OS (OS IA2)	167 events [redacted] Alpha boundary: 0.030	• If EFS IA1 not significant: triggered by EFS IA2. [redacted]
Final EFS (EFS FA) and third OS (OS IA3) interim analysis	185 events [redacted] Alpha boundary: 0.038	• If EFS IA2 not significant: triggered by EFS FA. [redacted]
Final OS analysis (OS FA)	-	[redacted]

* Estimated from the piecewise model described in Section 5.2

5.4 Analyses Timing Projections

The pCR analysis occured with a database lock on 16-Sep-2020.

Considering the actual enrollment, it will take

- Approximately 48 months when 148 events on Arms B and C (after start of 1:1:1 randomization) are observed for the first interim analysis (EFS IA1, OS IA1). This is

about 54 months from FPFV of the study. This analysis is triggered by the number of EFS events. In case of significant EFS, OS will also be tested at that time (OS IA1).

- Approximately 58 months when 167 EFS events on Arms B and C (after start of 1:1:1 randomization) are observed for the second interim analysis (EFS IA2, OS IA2). This is about 64 months from FPFV of the study. [redacted]
- Approximately 73 months when 185 EFS events on Arms B and C (after start of 1:1:1 randomization) are observed for the final EFS analysis (EFS FA, OSIA3). This is about 79 months from FPFV of the study. [redacted]

[redacted]

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. Evaluations (laboratory tests, pulse oximetry and vital signs) on the same date and time of the first dose of study treatment will be considered as baseline evaluations. Events (AEs) on the same date and time of the first dose of study treatment will not be considered as pre-treatment events.
  - 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 observations in the baseline period, then the latest non missing observation will be used as the baseline in the analyses. If multiple observations exist on the latest collection date (and time if collected), the record with the latest data entry date and time will be used. If multiple observations exist on the latest collection date (and time if collected) and data entry date and time, then the first observation is used as baseline, unless otherwise specified.
    - For PD-L1, non-missing is identified as those with quantifiable test result. After applying the rule above, if there are no records with a quantifiable test result, then select those with indeterminate result ("INDETERMINATE"). If there are no records with

indeterminate test result, then select those with unavailable result ("NOT EVALUABLE"). If there are no records with unavailable test result, then select those which are not reported or not available result (all other records).

- Post baseline period:
  - Neoadjuvant on-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of neoadjuvant study treatment (or with an onset date on or after the day of first dose of neoadjuvant study treatment if time is not collected or is missing). For participants who are off neoadjuvant study treatment, AEs will be included if event occurred within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of neoadjuvant 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.
  - Neoadjuvant 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 neoadjuvant study treatment. For participants who are off neoadjuvant study treatment, evaluations should be either within a safety window of 30 days (or 100 days depending on the analysis) after the last dose of neoadjuvant study treatment.
  - Adjuvant on-treatment AEs will be defined as AEs with an onset date and time on or after the date and time of the first dose of adjuvant systemic study treatment (or with an onset date on or after the day of first dose of adjuvant systemic study treatment if time is not collected or is missing). For participants who are off adjuvant study treatment, AEs will be included if event occurred within a safety window of 30 days after the last dose of adjuvant study treatment.

6.2 Treatment Regimens

The treatment group "as randomized" corresponds to the treatment group assigned by the Interactive Response Technology (IRT) system.

The treatment group "as treated" will be same as the treatment group "as randomized" by IRT unless a subject received the incorrect study treatment for the entire period of treatment, in which case the subject's treatment group "as treated" will be defined as the incorrect study treatment.

Unless otherwise specified, the safety analysis will be based on the treatment group "as treated".

Unless otherwise specified, the efficacy analysis will be based on the treatment group "as randomized".

The treatment arms are as follows:

- Arm A: nivolumab 3 mg/kg plus ipilimumab 1 mg/kg.
- Arm B: platinum doublet chemotherapy
- Arm C: nivolumab 360 mg plus platinum doublet chemotherapy

6.3 Populations for Analyses

- All Enrolled Participants: All participants who signed an informed consent form and were registered into the IRT.

- All Randomized Participants: All participants who were randomized to any treatment group in the study.
- All Treated Participants: All participants who received at least one dose of any study medication in neoadjuvant setting. This is the primary dataset for drug exposure and safety analysis for arm A.
- All Concurrently Randomized Participants in Arms B and C: All participants concurrently randomized on Arms B and C as of the 1:1:1 randomization. This will be the **primary analysis population** for efficacy.
- All Concurrently Randomized Participants in Arms A and B: All participants concurrently randomized on Arms A and B.
- All Treated Participants from the Concurrently Randomized Arms B and C: All participants concurrently randomized on Arms B and C as of the 1:1:1 randomization who received at least one dose of any study medication in the neoadjuvant setting. This will be the primary analysis population for drug exposure and safety for arms B and C.
- Tumor Tissue TMB evaluable subjects: All randomized subjects from the global study population with baseline evaluable tumor tissue TMB (non-missing numeric).

Concurrently randomized subjects from arms B and C (as of the 1:1:1 randomization) is considered at the site level basis, when the site switched to the revised protocol. In practice, this includes subjects randomized on the randomization lists from the 1:1:1 randomization (revised protocol 02) and the subsequent 1:1 randomization between B and C only (revised protocol 03).

Unless otherwise specified, 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 specified, analyses will be performed by treatment group (as randomized or as treated, depending on the analysis) for all concurrently randomized participants from Arms B and C. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings and in the consistency by randomization period analyses (Sections 7.3.7, 7.6.8 and 7.7.19).

Unless otherwise noted, discrete variables will be tabulated by the frequency and proportion of subjects falling into each category, grouped by treatment. Percentages given in these tables will be rounded to the first decimal and, therefore, may not always sum to 100%. Percentages less than 0.1 will be indicated as '< 0.1'. 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. Confidence intervals for binomial proportions will be derived using the Clopper-Pearson method[9].

Continuous variables will be summarized by treatment group using the mean, standard deviation, median, minimum, and maximum values and quartiles.

Time-to-event variables (e.g. time-to resolution, EFS) will be analyzed using the Kaplan-Meier technique. When specified, the median will be reported along with 95% CI using Brookmeyer and Crowley method[10] (using log-log transformation for constructing the confidence intervals[11]). Rates at fixed timepoints (e.g., OS at 12 months) will be derived from the Kaplan Meier estimate along with their corresponding log-log transformed confidence intervals[12].

Unless otherwise specified, the stratified hazard ratio between 2 treatment groups along with CI will be obtained by fitting a stratified Cox model with the treatment group variable as unique covariate. Stratification factors per IRT (PD-L1 expression (≥1% or <1%/not evaluable/indeterminate), disease stage (IB/II vs IIIA) and gender).

Unless otherwise specified, the stratified log-rank test will be performed to test the comparison between time to event distributions (OS and EFS). Stratification factors will be as described above.

The p-values from sensitivity analyses for efficacy endpoints, if presented, are for descriptive purpose only and not adjusted for multiplicity.

The conventions to be used for imputing missing and partial dates for analyses requiring dates are described in Section 8.

Note that in this document the terms "participant" and "subject" are used interchangeably. Terminology used in CSR will follow the BMS standard at the time of CSR.

Additional analyses by country or region may be conducted separately for country specific submissions.

7.1.1 Adverse Events, Serious Adverse Events, Multiple Events, Select Adverse Events, Other Events of Special Interest and Immune-Mediated Adverse Events

Drug-related AEs are those events with relationship to study drug "Related", as recorded on the CRF. If the relationship to study drug is missing, the AE will be considered as drug-related.

Serious adverse events consist of AEs deemed serious by the Investigator and flagged accordingly in the CRF and clinical database.

Adverse events leading to study drug discontinuation are AEs with action taken regarding study drug(s) = "Drug was discontinued". This option is selected when at least one agent from the regimen is discontinued.

Adverse events leading to dose delay are AEs with action taken regarding study drug(s) = "Drug was delayed".

Adverse events leading to dose reduction are AEs with action taken regarding study drug(s) = "Dose was reduced".

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), and the most recent version of the dictionary at the time of the database lock will be used. Adverse events results will be graded for severity using NCI Common Terminology Criteria for Adverse Events (CTCAE) and the version of the criteria specified in the protocol will be used (version 4).

In the AE summary tables, unless otherwise specified, subjects will be counted only once at the Preferred Term (PT), only once at the System Organ Class (SOC), and only once at subject level for the counting of total number of subjects with an AE. The AE tables will be sorted by the SOCs and then PTs. SOC will be ordered by descending frequency overall and then alphabetically. PTs will be ordered within SOC by descending frequency overall and then alphabetically. The sorting will be done based on the 'Any Grade' column of the experimental arm when arms are presented side-by-side.

Unless otherwise specified, the AE summary tables will be restricted to on-treatment events regardless of the causality.

Analyses that take into account the multiple occurrences of a given adverse event will be conducted (see Section 7.7.10). To prepare these analyses, the CRF data will be processed according to standard BMS algorithms[13] in order to collapse adverse event records into unique records based on the preferred term. These data will be presented as the rate per 100 person-years of exposure. These analyses will take into account all on-treatment events (allowing more than 1 event per subject) and the total exposure time. The person-year exposure will be computed as the sum over the subjects' neoadjuvant exposure expressed in years where the exposure time is defined as

- (Date of last dose of study treatment - date of first dose of study treatment + 31 days (or 101 days, depending on the analysis))/365.25, for subject who are off study treatment and were followed for at least 30 days (or 100 days, depending on the analysis) after last dose of study treatment.
- (Last known alive date - date of first dose of study treatment +1)/365.25, for subjects who are still on-treatment or who are off study treatment and were followed less than 30 days (or 100 days depending on the analysis) after last dose of study treatment.

7.1.1.1 Select Adverse Events

The select Adverse Events (select AEs) consist of a list of preferred terms grouped by specific category (e.g. pulmonary events, gastrointestinal events categories, etc.). AEs that may differ from or be more severe than AEs caused by non-immunotherapies and AEs whose early recognition and management may mitigate severe toxicity are included as select AEs. Categories of select AEs may include subcategories (e.g. adrenal disorders, diabetes, pituitary disorders, and thyroid disorders are subcategories of the endocrine event category).

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 time-to onset and time-to resolution are described in APPENDIX 1.

7.1.1.2 Other Events of Special Interest

Other events of special interest (OEOSI) 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.

7.1.1.3 Immune-Mediated Adverse Events

In order to further characterize AEs of special clinical interest, analysis of immune-mediated AEs (IMAE) will be conducted. IMAEs are specific 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 100 days of the last dose,
- 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.

7.1.2 Laboratory Tests

Clinical laboratory parameters (hematology, serum chemistry and electrolytes) will be evaluated.

Laboratory tests will be graded using the NCI Common Terminology Criteria, and the most recent version of the criteria at the time of the database lock will be used.

Clinical laboratory data will be analyzed using International System of Units (SI). Analyses will be repeated using US conventional units.

In the laboratory summary tables, unless otherwise specified, subjects will be counted only once for each lab parameter according to their worst on treatment CTC grade (worst being the highest CTC grade). The laboratory tables and listings will be sorted by laboratory category, laboratory subcategory and laboratory test code sequence number.

7.2 Study Conduct

Unless otherwise specified, analyses will be performed by treatment group as randomized for all concurrently randomized participants from Arms B and C. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings.

For analysis based on enrolled subjects, the analysis population will consist of the all enrolled population.

7.2.1 Accrual

Enrollment and randomization by country and site, and enrollment and randomization by month will be summarized and listed for all enrolled and randomized subjects.

7.2.2 Relevant Protocol Deviations

Unless otherwise specified, analyses will be performed by treatment group as randomized for all concurrently randomized participants from Arms B and C. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings.

Eligibility:

- Inadequate disease stage: presence of locally advanced unresectable (regardless of stage), stage IIIB or metastatic disease (stage IV) or stage IA disease.
- Subjects without measurable disease at baseline as per investigator.
- Subject with baseline ECOG performance status > 1.

On-study:

- Subjects receiving any concurrent anti-cancer therapy (chemotherapy, hormonal therapy, immunotherapy, radiation therapy, cancer related surgery (except definitive surgery), standard or investigational agents for treatment of cancer) outside of the protocol-specified neoadjuvant and adjuvant therapy (systemic and randiotherapy) while on study therapy (i.e. neoadjuvant or protocol adjuvant systemic treatment).
- Subjects whose "as treated" arm different than their as randomized arm (subjects who received the wrong treatment for the entire neoadjuvant treatment period, excluding the never treated)

7.3 Study Population

Unless otherwise specified, analyses will be performed by treatment group as randomized for all concurrently randomized participants from Arms B and C. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings.

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 on the all enrolled subjects population.

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

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 (including covid-19 reason). This analysis will be restricted to the all treated subjects population.

A by-subject listing for all treated subjects will be provided showing the subject's off treatment date along with the reason for going off treatment period. A by-subject listing for all enrolled subjects will also be provided, showing whether the subject was randomized/treated along with the reason for not being randomized/treated.

7.3.2 Demographics and Other Baseline Characteristics

The following demographic and baseline characteristics will be summarized and listed by treatment group as randomized:Age (descriptive statistics)

- Age (continuous)
- Age categorization (< 65, ≥ 65 and < 75, ≥ 75 and < 85, ≥ 85, ≥ 75, ≥ 65)
- Sex (male vs. female, CRF)
- Sex (male vs. female, IRT)
- Race (white, black , asian [asian Indian, Chinese, Japanese, asian other], American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Ethnicity (Hispanic/Latino, Not Hispanic/Latino, Unknown) - Required for US subjects only, but any available data will be presented
- Country by geographic region (North America, Europe, Asia, Rest of World)

The following baseline disease characteristics will be summarized by treatment group as randomized.

- Baseline ECOG performance status
- Baseline weight
- Tobacco use (Never Smoker, Current/Former, Unknown).
- Electronic Cigarette Use (Never Smoker, Current/Former, Unknown).
- Disease stage at study entry (CRF)
- Disease stage at study entry (IRT)
- Cell type (histology) at study entry (squamous cell carcinoma, non-squamous: adenocarcinoma, large cell carcinoma, broncho-alveolar carcinoma, other)
- Time from current NSCLC diagnosis to randomization
- Sites of diseases (all lesions) per BICR,
- Number of disease sites per subject (all lesions) per BICR
- Number of target lesions, non-target lesions and disease sites at baseline as per BICR

- Sum of the diameters of target lesions as baseline per BICR
- PD-L1 expression subgroups (clinical database, (<1%, >=1%, 1-49%, >=50%, inderterminate, not evaluable))
- PD-L1 expression subgroups (IRT, (<1%, >=1%/indeterminate/not evaluable)
- Tumor Tissue TMB subgroups (≥12.3 mut/MB, < 12.3 mut/MB, not evaluable)

Summary table (cross-tabulation) by treatment group for stratification factor will be provided to show any discrepancies between what was reported through IRT vs. CRF/Clinical database at baseline. This summary will be performed based on all randomized subjects.

- PD-L1 status (IRT vs. clinical database)
- Disease stage (IRT vs CRF data)
- Gender (IRT vs CRF data)
-

A listing of randomization scheme presenting randomized treatment group and as treated treatment group will be provided for all randomized subjects.

7.3.3 Medical History

A by-subject listing of general medical history for all randomized subjects will be provided.

7.3.4 Prior Therapy

- Prior/current non-study medication classified by anatomic and therapeutic classes.

Agents and medication will be reported using the generic name. A listing by subject will also be provided.

7.3.5 Physical Examinations

Subjects with abnormal baseline physical examination will be listed by subject.

7.3.6 Baseline Physical Measurements

Baseline physical measurements will be listed by subject.

7.3.7 Consistency of Demographics and Baseline Characteristics by Randomization Period

Consistency of population amongs the different randomization periods will be examined by summarizing the characteristics listed in Section 7.3.2 by treatment arm as randomized in the the different randomization period: before revised protocol 02 (1:1 A vs B), under revised protocol 02 (1:1:1 A vs B vs C) and after revised protocol 03 (1:1 B vs C).

Concurrent randomization is considered at the site level basis, when the site switched to the revised protcol. In practice, this includes subjects randomized on the randomization lists from the 1:1:1 randomization (revised protocol 02) and the subsequent 1:1 randomization between B and C only (revised protocol 03).

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 and for all treated participants from concurrently randomized Arms B and C, unless otherwise specified. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings.

7.4.1 Administration of Study Therapy

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

- Number of neoadjuvant doses received by drug
- Cumulative dose by drug in neoadjuvant
- Relative dose intensity (%) by drug in neoadjuvant using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; ≥ 110%
- Number of subjects who received protocol specified adjuvant systemic therapy number of adjuvant doses received by drug for subjects who received adjuvant therapy.
- The frequency of subjects receiving carboplatin instead of cisplatin in the regimens other than carboplatin-paclitaxel will be reported and the reason for not using cisplatin for regimens other that carboplatin-paclitaxel will be summarized.

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.

- Number of subjects who received adjuvant radiotherapy, including radiotherapy type, number of doses and total cumulative dose.

Table 7.4.1-1 to Table 7.4.1-6 summarize the key parameters used to calculate dosing data.

Table 7.4.1-1: Study Therapy Parameter Definitions- Nivolumab and Ipilimumab

	Nivolumab	Nivolumab	Ipilimumab
Dosing schedule per protocol	3 mg/kg every 2 weeks	360 mg every 3 weeks	1 mg/kg on first cycle
Dose	*Dose (mg/kg)* is defined as Total Dose administered (mg)/Most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.	*Dose (mg)* is defined as Total Dose administered (mg) at each dosing date as collected on the CRF.	*Dose (mg/kg)* is defined as Total Dose administered (mg)/Most recent weight (kg). Dose administered in mg at each dosing date and weight are collected on the CRF.
Cumulative Dose	*Cum dose (mg/kg) is* sum of the doses (mg/kg) administered to a subject.	*Cum dose (mg)* is the sum of the doses (mg) administered to a subject.	*Cum dose (mg/kg) is* sum of the doses (mg/kg) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/kg)/[(Last Nivolumab dose date - Nivolumab start dose date + 14) x 3/14] x 100	Cum dose (mg)/[(Last Nivolumab dose date - Nivolumab start dose date + 21) x 360/21] x 100	Cum dose (mg/kg) x 100

Table 7.4.1-2: Study Therapy Parameter Definitions - Regimen 1: Vinorelbine/Cisplatin

	Vinorelbine	Cisplatin
Dosing schedule per protocol	25 mg/m^2 or 30 mg/m^2 on Day1 and Day8 of a 3 week cycle.	75mg/ m² on Day 1 of a 3 week cycle
Dose	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.
Cumulative Dose	*Cum dose (mg/ m²) is* sum of the doses (mg/ m^2) administered to a subject.	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/m²)/[(First Vinorelbine dose date in the last cycle - Vinorelbine Start dose date + 21) x 50/21] x 100 or Cum dose (mg/m²)/[(First Vinorelbine dose date in the last cycle - Vinorelbine Start dose date + 21) x 60/21] x 100	Cum dose (mg/m²)/[(Last Cisplatin dose date - Start Cisplatin dose date + 21) x 75/21] x 100

Table 7.4.1-3: Study Therapy Parameter Definitions - Regimen 2: Docetaxel/Cisplatin

	Docetaxel	Cisplatin
Dosing schedule per protocol	60 mg/m^2 or 75 mg/m^2 on Day1 of a 3 week cycle.	75mg/ m² on Day 1 of a 3 week cycle
Dose	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.
Cumulative Dose	*Cum dose (mg/ m²) is* sum of the doses (mg/ m^2) administered to a subject.	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/m²)/[(First Docetaxel dose date in the last cycle - Docetaxel Start dose date + 21) x 60/21] x 100 or Cum dose	Cum dose (mg/m²)/[(Last Cisplatin dose date - Start Cisplatin dose date + 21) x 75/21] x 100

Table 7.4.1-3: Study Therapy Parameter Definitions - Regimen 2: Docetaxel/Cisplatin

	Docetaxel	Cisplatin
	(mg/m²)/[(First Docetaxel dose date in the last cycle - Docetaxel Start dose date + 21) x 75/21] x 100	

Table 7.4.1-4: Study Therapy Parameter Definitions - Regimen 3: Gemcitabine/Cisplatin

	Gemcitabine	Cisplatin
Dosing schedule per protocol	1250 mg/m^2 or 1000 mg/m^2 on Day1 and Day8 of a 3 week cycle.	75mg/ m² on Day 1 of a 3 week cycle
Dose	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.
Cumulative Dose	*Cum dose (mg/ m²) is* sum of the doses (mg/ m^2) administered to a subject.	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/m²)/[(First Gemcitabine dose date in the last cycle - Gemcitabine Start dose date + 21) x 2500/21] x 100 or Cum dose (mg/m²)/[(First Gemcitabine dose date in the last cycle - Gemcitabine Start dose date + 21) x 2000/21] x 100	Cum dose (mg/m²)/[(Last Cisplatin dose date - Start Cisplatin dose date + 21) x 75/21] x 100

Table 7.4.1-5: Study Therapy Parameter Definitions - Regimen 4: Pemetrexed/Cisplatin

	Pemetrexed	Cisplatin
Dosing schedule per protocol	500 mg/ m² every 3 weeks	75mg/ m² every 3 weeks

Table 7.4.1-5: Study Therapy Parameter Definitions - Regimen 4: Pemetrexed/Cisplatin

	Pemetrexed	Cisplatin
Dose	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (mg/m²)* is defined as Total Dose administered (mg)/Most recent BSA. Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.
Cumulative Dose	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/m²)/[(Last Pemetrexed dose date - Pemetrexed Start dose date + 21) x 500/21] x 100	Cum dose (mg/m²)/[(Last Cisplatin dose date - Cisplatin Start dose date + 21) x 75/21] x 100

Table 7.4.1-6: Study Therapy Parameter Definitions - Regimen 5: Paclitaxel/Carboplatin

	Paclitaxel	Carboplatin
Dosing schedule per protocol	200 mg/ m^2 or 175 mg/ m^2 every 3 weeks	AUC 5 or 6 every 3 weeks
Dose	*Dose (mg/m²)* is defined as Total Dose administered in mg at each dosing date is collected on the CRF and BSA is derived from most recent weight and baseline height also collected on the CRF.	*Dose (AUC)* is defined as Total Dose administered (mg)/(creatinine clearance +25). Dose administered in mg at each dosing date is collected on the CRF and creatinine clearance derived from the CRF data and capped at 125 mL/min
Cumulative Dose	*Cum dose (mg/m²) is* sum of the doses (mg/m²) administered to a subject.	*Cum dose (AUC) is* sum of the doses (AUC) administered to a subject.
Relative dose intensity (%)	Cum dose (mg/m²)/[(Last paclitaxel dose date - paclitaxel Start dose date + 21) x 200/21] x 100 or Cum dose (mg/m²)/[(Last paclitaxel dose date - paclitaxel Start dose date + 21) x 175/21] x 100	Cum dose (AUC)/[(Last dose date of Carbo - Start dose date of Carbo + 21) x 6/21] x 100 or Cum dose (AUC)/[(Last dose date of Carbo - Start dose date of Carbo + 21) x 5/21] x 100

Where the creatinine clearance will be calculated using Cockroft-Gault formula, defined as:

$$\mathrm{CrCL(ml/min)} = \frac{(140 - \mathrm{age(in\ years)}) * \mathrm{weight(in\ kg)}}{72 * \mathrm{serum\ creatinine(in\ mg/dL)}}$$

for males and

$$\text{CrCL(ml/min)} = \frac{(140 - \text{age(in years)}) * \text{weight(in kg)}}{72 * \text{serumcreatinine(in mg/dL)}} * 0.85$$

for females. The most recent weight will be used. If the computed creatinine clearance is more than 125 ml/min, then the creatinine clearance value should be capped at 125ml/min for dose exposure computations.

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose Delay/Omission

Each study medication infusion may be delayed. A dose will be considered as actually delayed if the delay is exceeding 3 days (i.e., greater than or equal to 4 days from scheduled dosing date) for study medication. In case of reported omission of a dose between 2 doses, this will be taken into account in the derivation and will not count as a delay unless there is a delay in addition to the omission.Reason for dose delay/omission will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment group and by drug.

- Number of subjects with at least one dose delayed, the number of dose delays per subject, the reason for dose delay (including covid related reason) and the length of dose delay.
- Number of subjects with reported dose omission, reason for omission.
- The listing of dosing will include the dose modifications, including dose modifications for adjuvant systemic therapy. Dose modifications for adjuvant therapy will only include the reported modifications, there will be no derivation.

7.4.2.2 Infusion Interruptions and Rate Changes

Each study drug infusion can be interrupted and/or the IV infusion rate can be reduced. This information will be retrieved from CRF dosing pages

The following parameters will be summarized by treatment group and study drug:

- Number of subjects with at least one dose infusion interruption, the reason for interruption, and the number of infusion interruptions per subject.
- Number of subjects with at least one IV infusion rate reduction, the reason for reduction and the number of infusion with IV rate reduction per subject.

7.4.2.3 Dose Reductions

There will be no dose reductions of nivolumab and ipilimumab allowed. Dose of platinum doublet chemotherapy (Arms B and C) may be modified for toxicity. Dose levels of platinum doublet chemotherapy (Arms B and C) are defined in the protocol as follows:

Table 7.4.2.3-1: Dose Modifications of Chemotherapeutic Agents (Arms B and C)

Dose Level	Vinorelbine	Docetaxel	Gemcitabine	Pemetrexed	Cisplatin	Carboplatin	Paclitaxel
Starting dose	25 mg/m² or 30 mg/m²	60 mg/m² or 75 mg/m²	1000 mg/m² or 1250 mg/m²	500 mg/m²	75 mg/m²	AUC 5 or 6	175 or 200 mg/m²
First dose reduction	75% of strating dose	75% of starting dose	75% of starting dose	75% of starting dose	75% of starting dose	AUC 4 or 5	150 mg/m²
Second dose reduction	50% of starting dose	50% of starting dose	50% of starting dose	50% of starting dose	50% of starting dose	AUC 3 or 4	100 mg/m²
Third dose reduction	Discontinue	Discontinue	Discontinue	Discontinue	Discontinue	Discontinue	Discontinue

For any cycle, it will be defined as a dose reduction if the observed dose level (based on calculated administered dose) is below protocol specified dose level. Dose ranges for dose levels of platinum doublet chemotherapy are defined in Table 7.4.2.3-2.

Table 7.4.2.3-2: Calculated Dose Ranges and Related Dose Levels

Dose Level	Dose Range						
	Vinorelbine (mg/m²)	Docetaxel (mg/m²)	Gemcitabine (mg/m²)	Pemetrexed (mg/m²)	Cisplatin (mg/m²)	Carboplatin (AUC)	Paclitaxel (mg/m²)
Level 0	≥21.875 or ≥26.25	≥ 52.5 or ≥65.625	≥875 or ≥1093.75	≥437.5	≥65.625	≥4.5 or ≥5.5	≥153.125 or ≥175
Level -1	<21,875 and ≥ 15.625 or <26.25 and ≥ 18.75	<52.5 and ≥37.5 or < 65.625 and ≥46.875	<875 and ≥625 or <1093.75 and ≥781.25	<437.5 and ≥312.5	<65.625 and ≥46.875	<5.5 and ≥4.5 or <4.5 and ≥3.5	<175 and ≥125
Level -2	<15.625 or <18.75	<37.5 or < 46.875	<625 or < 781.25	<312.5	<46.875	<4.5 or <3.5	<125

The reason for dose reduction as reported by the investigator will be tabulated for all instances of dose reduction based on the Dose Change CRF page. A category 'Unknown' will be defined for all reductions with no reason reported by the investigator.

Chemotherapy dose reductions are permanent; once the dose of any chemotherapy agent is reduced, it may not be re-escalated in subsequent cycles.

The following will be summarized for chemotherapeutic agent arm only:

Number and percentage of subjects with at least one dose reduction and reason of the dose reduction, number and percentage of subjects with a dose reduction to dose level -1, number and percentage of subjects with a dose reduction to dose level -2.

7.4.3 Concomitant Medications

Concomitant medications, defined as medications other than study medications which are taken at any time on-treatment (i.e. on or after the first day of study therapy and within 100 days following the last dose of neoadjuvant study therapy or within 30 days following the last dose of adjuvant therapy, whichever is longest), will be coded using the WHO Drug Dictionary.

The following summary tables by treatment group will be provided:

- Concomitant medications (subjects with any concomitant medication, subjects by medication class and generic term).

Prior medications, defined as non-study medications with a start date before consent date, and current medications, defined as non-study medications with a start date before the first date of study medication and stop date after consent date, will be coded using the WHO Drug Dictionary.

The following summary table will be provided:

- Prior/current medications (subjects with any prior/current medication, subjects by medication class and generic term)

By-subject listings will accompany the tables.

7.4.3.1 Immune Modulating Medication

Immune modulating concomitant medications are medications entered on an immune modulating medication form or available from the most current pre-defined list of immune modulating medications. The list of anatomic class, therapeutic class and generic name used for the selection at the time of the database lock will be provided.

The percentage of subjects who received immune modulating concomitant medication for

- management of adverse event
- premedication
- other use
- any use
- management of drug-related select adverse event (any grade, grade 3-5) by select AE category/ subcategory
- management of IMAEs (any grade, grade 3-5) by IMAE category

will be reported separately for each treatment group (percentages of treated subjects 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 for each treatment group:

- The total immune modulating medication treatment duration (excluding overlaps), duration of high dose of corticosteroid, initial dose of corticosteroid, and tapering duration (summary statistics)

Duration represents the total duration the subject received the concomitant medication of interest. If the subject took the medication periodically, then DURATION in the summation of all use. Initial dose represents the dose of the concomitant medication of interest received at the start of the event. In the case multiple medications started on the same date, the highest equivalent dose is chosen and converted to mg/kg by dividing by the subject's recent weight.

These analyses, except the ones related to IMAEs will be conducted using the 30-day safety window. The analyses related to IMAEs will be conducted using the 100-day safety window.

7.4.3.2 Subsequent Cancer Therapy

Subsequent therapies are defined as Cancer therapies started on or after the first study drug dose or date of randomization if the subject is no treated, outside of the protocol defined adjuvant therapy (systemic and radiotherapy).

The following information pertaining to subsequent therapies will be summarized by treatment arm, as randomized:

- Number and percentage of subjects receiving subsequent therapies including:
- Subsequent systemic therapy by drug name
- Subsequent disease related surgery
- Subsequent radiotherapy for treatment of tumors

A by-subject listing of subsequent cancer therapy will also be produced for randomized subjects.

7.5 Definitive Surgery

Unless otherwise specified, analyses will be performed by treatment group as randomized for all concurrently randomized participants from Arms B and C. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings.

The following parameters will be summarized:

- Disease Stage Prior to Surgery
- Subjects with clinical downstaging (lower stage prior to surgery vs baseline)
- Subjects with surgery
-

Subjects without surgery:

- Reason for cancelled surgery

Subjects with surgery:

- Delayed surgery (>6 weeks post last neoadjuvant dose), reason for delay as reported in CRF

- Duration of delay (number of weeks between last neoadjuvant dose and surgery date exceeding 6 weeks), descriptive statistics and categories: 1-2 weeks, 3-4 weeks, 5-6 weeks, >6 weeks delay)
- Duration of surgery
- Length of hospitalization for definitive surgery
- Method of surgery (Minimally invasive-thoracoscopic/robotic, Thoracotomy, Minimally innvasive to thoracotomy)
- Type of Surgery (Pneumonectomy, Lobectomy, Sleeve Lobectomy, Bilobectomy, Other)
- Surgery Outcome (R0, R1, R2, unknown)

Safety related to surgery analyses are described in section 7.7.1

7.6 Efficacy

Unless otherwise specified, analyses will be performed by treatment group as randomized for all concurrently randomized participants from Arms B and C. Descriptive analyses will also be produced for Arm A. Participants in Arm B randomized in the initial protocol will only be reported in listings and in analyses based on the All Concurrently Randomized Participants in Arms A and B population.

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

- PD-L1 expression (≥1% or <1%/not evaluable/indeterminate)
- Disease stage (IB/II vs IIIA)
- Gender

Alpha (α) for the confidence intervals (CIs) for hazard ratios, odds ratios or difference of rates will be the same as nominal significance level for hypothesis testing. CIs for endpoints not tested will be at the two-sided 95% level. All p-values reported will be two-sided. P-values will be rounded to the fourth decimal place. Point estimates and confidence bounds for efficacy variables will be rounded to the second decimal place.

7.6.1 Type I Error Control

The overall alpha will be controlled using the following procedure. The overall alpha is primarily allocated to the two primary endpoints: 1% for pCR and 4% for EFS.

- The primary endpoint pCR will be tested at 1% alpha.
- If pCR is not significant, the primary endpoint EFS will be tested at 4%
- If pCR is significant, the 1% alpha will be re-allocated to the EFS primary endpoint which will be tested at 5% alpha level
- If EFS is significant, OS will be tested at the same level as EFS

EFS and OS (if EFS is significant) will be tested at planned interim and final analyses. Stopping boundaries will be calculated for each endpoint according to the observed number of events by Lan-DeMets alpha spending function with O'Brien-Fleming boundaries corresponding to an overall alpha of 4% or 5%. Given EFS and OS endpoints are tested using group sequential

approach, overall hierarchical testing approach will be used where each endpoint will have its own specific Lan-DeMets alpha spending function with O'Brien-Fleming boundaries[8]. Also refer to Sections 5.2 and 5.3.

If the p-value crosses the boundary at the interim analysis (EFS or OS), the p-value from the interim stratified log-rank test will be considered the final analysis result for the study.

The secondary endpoints of Major Pathologic Response and Time to Death or Distant Metastases will be analyzed descriptively without hypothesis testing.

7.6.2 Analysis of Pathological Complete Response

7.6.2.1 Primary pCR Analysis

Formal analysis of pCR will occur after the 350 randomized participants in arms B and C from start of 1:1:1 randomization have an opportunity for surgery.

At pCR analysis, the primary analysis population is the concurrently randomization participants in arms B and C. PCR rate will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method.

The numerator is based on randomized participants achieving pCR in both tumor and lymph nodes, as assessed by independent pathological review (BIPR). The denominator is based on All Concurrently Randomized Participants in Arms B and C. Subjects who are no longer eligible for surgery, or who are on alternative anti-cancer therapy before surgery, or who discontinue before surgery or for whom pCR results are not available are all counted as non-responders.

pCR will be compared between concurrent arms B and C by the stratified Cochran Mantel-Haenszel (CMH) test using a 2-sided, 1% alpha level.

An estimate of the difference in pCR rates between the treatment groups along with the corresponding two sided 99% CI will also be computed using the following Cochran-Mantel-Haenszel (CMH) method of weighting, adjusting for stratification factors[14]. A two sided 99% CI for odds ratio of pCR between the treatment groups will also be computed.

Estimate of the difference in pCR between arms A and B (in the All Concurrently Randomized Participants in Arms A and B population) and odds ratio will also be provided together with corresponding 95% CI using the same methodology.

The analysis will be conducted by an independant statistician external to BMS and reviewed by the DMC. At the time of pCR analysis, EFS descriptive analyses (by investigator and by BICR) will be produced in the DMC closed report and might be shared with regulatory authorities. The communication of results will be tightly controlled and pre-specified in the DMC charter to maintain trial integrity.

7.6.2.2 Supportive Analyses of pCR

pCR sensitivity analyses will be performed with the following consideration:

- pCR analysis will be repeated for response evaluable subjects, where response evaluable subjects are subjects who had definitive surgery, and didn't start alternative anti-cancer therapy

before surgery and pathologic samples results at surgery are evaluable. No p-value will be generated.

- pCR using stratification factors as obtained from the baseline CRF pages or database (instead of IRT). This analysis will be performed only if the stratification variable/factor at randomization (as per IRT) and baseline are discordant for at least 10% of randomized subjects. Stratified Cochran Mantel-Haenszel (CMH) p-value will be generated.
- Considering the small number of subjects that may be impacted by the Covid-19 situation in terms of pathology assessment, no sensitivity analysis is currently planned. The potential surgeries delays or cancellation due to Covid-19 related issues will be reported based on the surgery listing.
-

7.6.2.3 Subset Analyses of pCR

The influence of baseline and demographic characteristics on the treatment effect will be explored via exploratory subset analysis. BIPR assessment of pCR will be summarized for the following subgroups:

- Age category
  a) <65,
  b) >=65 and <75
  c) >=75 and <85
  d) >=85
  e) >=75
  f) >=65
- Sex (male, female), per IRT and per CRF
- Race (white, black, Asian, other)
- Region (North America, Europe, Asia, Rest of World)
- Baseline ECOG Performance Status (0, 1, >1)
- Tobacco use (current/former, never smoked, unknown)
- Disease stage (IB/II vs IIIA) per IRT and per CRF
- Baseline histology (squamous, non-squamous)
- PD-L1 subgroups (<1%, ≥1%, 1-49%, ≥50%, indeterminate, not evaluable)
- Tumor Tissue TMB Evaluable (≥12.3 mut/MB, < 12.3 mut/MB, Overall)
- Tumor Tissue TMB Not Evaluable
- Type of platinum therapy (cisplatin, carboplatin, subjects switching from cisplatin to carboplatin).
- Type of chemotherapy regimen in arm B (available in arm C (Gemcitabine-Cisplatin, Pemetrexed-Cisplatin, Paclitaxel-Carboplatin, not available in arm C (Vinorelbine-Cisplatin, Docetaxel-Cisplatin)) , based on first neoadjuvant cycle.

A forest plot of treatment effect on pCR per BIPR in the above subgroups will be produced. The un-weighted differences in pCR between concurrent arms B and C and corresponding 95% two-sided CI using the method of Newcombe, will be provided.

The analysis comparing treatment (i.e., pCR difference) will be conducted if the number of subjects in the subgroup category is more than 10.

7.6.2.4 Major Pathological Response Rate

MPR rate in concurrently randomized participants in arms B and C will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method. An estimate of the difference and odds ratio in MPR rates between concurrent arms B and C and corresponding 95% CI will be calculated using CMH methodology and adjusted by stratification factors.

Estimate of the difference in MPR between arms A and B (in the All Concurrently Randomized Participants in Arms A and B population) and odds ratio will also be provided together with corresponding 95% CI.

Subset analyses by PDL1 status will be performed (PD-L1<1% PD-L1 ≥ 1%, PD-L1 1-49%, PD-L1 ≥ 50%, not evaluable/indeterminate) and by Tumor TMB (≥12.3 mut/MB, < 12.3 mut/MB, Overall, not evaluable).

7.6.2.5 Additional Pathological Related Analyses

Descriptive analyses of pCR and MPR, % tumor area with viable tumor cells in the tumor region and in lymph nodes separately will be provided.

7.6.2.6 Clinical Response Rate

Clinical response rate (cRR) by BICR will be summarized by treatment arm. cRR is defined as proportion of randomized participants whose radiologic response at the last scan prior to definitive surgery is either a complete response or partial response per RECIST 1.1 criteria by BICR. The response does not require confirmation. Response rates and their corresponding 95% exact CI will be calculated by Clopper-Pearson method presented for each randomized arm.

Clinical response rate by investigator will be reported similarly.

Subset analyses of cRR by BICR by PDL1 status will be performed (PD-L1<1% PD-L1 ≥ 1%, PD-L1 1-49%, PD-L1 ≥ 50%, not evaluable/indeterminate) and by Tumor TMB (≥12.3 mut/MB, < 12.3 mut/MB, Overall, not evaluable).

7.6.3 Analysis of Event Free Survival

7.6.3.1 Primary Event Free Survival

One of the primary objectives of the study is to compare the event-free survival (based on BICR assessments) between treatment groups in all concurrently randomized participants in Arms B and C.

The primary definition of EFS, censoring for subsequent anticancer therapy, will be used in this analysis.

EFS will be compared between the treatment groups (concurrent B and C) at the interim and final analyses, using stratified log-rank test, with stratification factors as per IRT, two-sided p-value will also be reported. A Lan DeMets α-spending function with O'Brien and Fleming type of boundary 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).

EFS will be estimated using the Kaplan Meier techniques and will be displayed graphically. A two-sided 95% CI for median EFS in each treatment group will be computed via the log-log transformation method. EFS rates at fixed time points (e.g. 6, 12 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood[15] formula for variance derivation and on log-log transformation applied on the survivor function[16].

These EFS analyses will also be conducted for Arm A, including Kaplan Meier curve, median and EFS rates with 95% CI and stratified HR between Arm A and Arm B (in the All Concurrently Randomized Participants in Arms A and B population) with 95% HR.

Analyses of EFS will also be conducted based on the secondary definition of EFS (not censoring for subsequent therapies). These analyses will be the same as those specified above.

The source of EFS event (progression precluding surgery, progression, recurrence (locoregional, distant) or death) will be summarized by treatment group. The status of subjects who are censored (as per primary definition of EFS) in the EFS KM analysis will be tabulated for each treatment group including the following categories:

- On-study (on-neoadjuvant treatment, on-adjuvant 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
- A by-subject listing will be presented including treatment group, EFS duration under the primary definition, EFS duration on the secondary definition, whether the subject was censored under the primary definition, and if censored, the reason, and whether the subject was censored under the secondary definition, and if censored, the reason.

A by-subject listing of lesion evaluations per BICR will be presented.

7.6.3.2 Supportive Analyses of Event-Free Survival

The following sensitivity analyses will be conducted in the concurrently randomized subjects in arms B and C for the primary definition. The p-values from sensitivity analyses for efficacy endpoints, are for descriptive purpose only and not adjusted for multiplicity.

- Delayed effect of immunotherapy interventions may cause a late separation in the OS KM curves and non-proportional hazards.

- EFS will be compared between treatment groups via a 2-sided max-combo test. The max-combo test statistic is the maximum of 4 diffrerent Fleming-Harrington family weighted log-rank test statistics. Zm = max (FH (0, 0), FH (0,1), FH (1,0), F(1,1)), where FH(ρ,γ) are the test statistics from the Fleming-Harrington family of test statistics. FH (0, 0) corresponds to the log-rank test, while FH (0, 1) is more sensitive to late-difference alternatives, FH(1,0) is more sensitive to early difference with decreasing treatment effect and FH(1,1) uses weigths at the median.
- To examine the assumption of proportional hazards in the Cox regression model, in addition to treatment, a time-dependent variable defined by treatment by time interaction will be added into the model. A two-sided Wald Chi-square p-value of less than 0.1 may indicate a potential non constant treatment effect. In such case, the following analysis will be conducted:
  - The estimates of the EFS hazard ratios will be estimated in 2 periods. The periods will be defined by a cut off point. The cut off point will be calculated using a stratified time-dependent Cox model with effects for treatment and period-by-treatment interaction. The cut off point will be estimated using a grid of possible cut off points and obtained by maximizing the partial log likelihood. Ties will be handled using the exact method. A two-sided 95% CI for the hazard ratio's will also be presented. Visual interpretation of the curves may lead to additional analyses with several cut off points.

- 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, will 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:
  - Histology (Squamous, Non-squamous)
  - Age categorization (< 65, ≥ 65)
  - ECOG (0, ≥ 1)
  - Race (White, Black, Asian, Other)
- 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 95% CIs and p-value.
- The primary EFS based on BICR assessments analysis will be repeated using secondary EFS definition which accounts for the tumor scans post subsequent therapies for the primary efficacy population. Stratified log-rank test p-value will be generated.
- EFS based on BICR assessments, using stratification factors as obtained from the baseline CRF pages or database (instead of IRT). This analysis will be performed only if the stratification variable/factor at randomization (as per IRT) and baseline are discordant for at least 10% of randomized subjects. Stratified log-rank test p-value will be generated.
- EFS based on BICR assessments accounting for missing tumor assessment prior to EFS event (progression/recurrence or death). This analysis will be performed only if at least 10% of events have missing prior tumor assessment within the primary efficacy population. It will apply the following restriction to the primary definition: If the elapsed time between the EFS event and the last assessment immediately prior to the event is two or more missed visits, the

subject's EFS will be censored at his/her last tumor assessment prior to the EFS event. Stratified log-rank test p-value will be generated. This analyses may account for potential missed assessments due to the Covid-19 situation.

- EFS based on BICR assessments accounting for site reported pathology results. In case of site reported pathology recurrence on the CRF at an earlier date than the BICR event date, an event will be assigned at the pathology site reported date. No p-value will be generated.
- EFS based on BICR assessments accounting for BICR assessed progressions occuring before surgery but not precluding surgery. This analysis will use the primary definition of EFS but will in addition count an event for BICR assessed RECIST 1.1 progression before surgery, that would not preclude surgery.
- EFS based on investigator assessments. The hazard ratio associated with treatment and median EFS will be presented along with the associated two-sided 95% CIs. Kaplan-Meier plot will be produced. It is to be noted that per CRF instruction, the investigator will not consider a second primary cancer as a recurrence/progression. While if such lesions are present on tumor assessment at the BICR, they will be considered as new lesions, since the BICR does not have access to biopsy results. EFS by investigator is defined the same way as for EFS by BICR, except that both pathology and imaging recurrences reported in the CRF are taken into account as event and censoring will occur at the time of last tumor assessment prior to (or at the date of) second primary cancer. No p-value will be generated.
- EFS based on BICR assessments using an un-stratified Cox model. Un-stratified log-rank test p-value will be generated
- EFS in treated subjects from concurrently randomized arms B and C using treatment group "as treated" if more than 10% randomized subjects in any treatment group were never treated or treated differently than randomized among corresponding analysis population. Stratified log-rank test p-value will be generated.
- EFS analysis for participants with no relevant deviation. This analysis will be conducted only if there are more than 10% participants with relevant protocol deviations. Stratified log-rank test p-value will be generated.
- In order to assess the potential impact of the change in tumor assessment scheduled on the longer term and potential missing assessments, EFS (by BICR) will also be analyzed based on interval censoring method. The SAS PROC ICPHREG will be used to fit the proportional hazards regression models using interval censoring approach, with treatment arm as the only covariate in the model. Hazard ratio with 2-sided 95% and 100-alpha confidence intervals will be produced. The time period between time1 and time2 will be the interval during which the EFS event occurred. Time1 and time2 will be setup as follows:
  - For subjects who had EFS event per EFS primary definition, time1 is the time from randomization date to the last tumor assessment date prior to EFS event date, and time2 is time to EFS event (EFS duration).
  - For subjects who are censored for EFS, time1 is time to EFS event (EFS duration), and time2 is infinite time (missing in SAS).
- In order to assess the potential variability introduced by the optional adjuvant chemotherapy:
  - Baseline demographics and disease characterictics, characteristics at the time of surgery will be tabulated by adjuvant therapy status.

  - A Cox regression model with treatment and an additional time-dependant covariate as indicator of the start of adjuvant therapy will used.
  - Additional sensitivity analyses may be performed
- Given the small number of subjects expected to be impacted by a death due to Covid-19 infection, no specifc sensitivity analysis is currently planned. Subjects with death (potentially) associated with Covid-19 infection will be reported based on the reason for death in the death listing.

7.6.3.3 Subset Analyses of EFS

The influence of baseline and demographic characteristics on the treatment effect will be explored via subset analyses for the factors specified in section 7.6.2.3.

A forest plot of the EFS based on BICR assessments, unstratified hazard ratios (HR) along with two-sided 95% CIs will be produced for each level of the subgroups listed in section 7.6.2.3. If subset category has less than 10 subjects per treatment group, HR will not be computed/displayed. Median and 95% CI will be provided.

In addition, for gender, baseline disease stage, histology, PD-L1 and tumor TMB subsets, Kaplan Meier Curves will be generated.

Kaplan Meier curves, medians and 95% confidence interval will be generated for EFS based on BICR assessments by PD-L1 ($\geq$ 1%, <1%, 1%-49%, $\geq$ 50%) for Arm A.

7.6.3.4 EFS Analyses by pCR and MPR Status

EFS (based on BICR assessments, primary definition) Kaplan-Meier curves will be generated by pCR and by MPR status. These analyses will be landmarked at the time of surgery and will be limited to subjects with pCR or MPR status available. Median and 95% CI will be provided. HR and 95% CI for concurrently randomized subjects in arms B and C will be provided by pCR and by MPR status, as well as HR of pCR/MPR vs no pCR/MPR by treatment arm. If subset category has less than 10 subjects per treatment group, HR will not be computed/displayed.

In addition, EFS (based on BICR assessments, primary definition) Kaplan-Meier curves will be generated by pCR and by MPR status, without landmark. These analyses will include all randomized subjects concurrently ranomdized to Arms B and C. Median and 95% CI will be provided. HR and 95% CI for concurrently randomized subjects in arms B and C will be provided by pCR and by MPR status, as well as HR of pCR/MPR vs no pCR/MPR by treatment arm. If subset category has less than 10 subjects per treatment group, HR will not be computed/displayed.

7.6.3.5 Current Status of EFS

Time from last censoring point to cutoff date in months will be summarized by treatment group and overall for randomized subjects. Subjects who have a EFS event will be considered as current for this analysis. The secondary definition of EFS (by BICR) will be used for this summary.

7.6.4 Analysis of Overall Survival

7.6.4.1 OS Analyses

OS will be hierarchically tested if EFS is significant. Details are provided in section 5.3.

OS will be compared between the treatment groups (concurrent B and C) at the interim and final analyses, using stratified log-rank test, with stratification factors as per IRT, two-sided p-value will also be reported. A Lan DeMets α-spending function with O'Brien and Fleming type of boundary 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).

OS will be estimated using the Kaplan Meier techniques and will be displayed graphically. A two-sided 95% CI for median EFS in each treatment group will be computed via the log-log transformation method. EFS rates at fixed time points (e.g. 6, 12 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

These analyses OS analyses will also be conducted for Arm A, including Kaplan Meier curve, median and OS rates with 95% CI and stratified HR between Arm A and Arm B (in the All Concurrently Randomized Participants in Arms A and B population) with 95% HR.

The number of participants who are censored in the OS KM analysis and their status will be tabulated using following categories:

- Still on treatment (No recurrence/progression, recurrence/progression)
- In follow-up
- Off study
  - lost to follow-up
  - participant withdrew consent
  - other

7.6.4.2 Supportive Analyses for OS

The following sensitivity analyses will be conducted in the concurrently randomized subjects in arms B and C. The p-values from sensitivity analyses for efficacy endpoints, if presented, are for descriptive purpose only and not adjusted for multiplicity.

- 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, will 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:
  - Histology (Squamous, Non-squamous)
  - Age categorization (< 65, ≥ 65)
  - ECOG (0, ≥ 1)
  - Race (White, Black, Asian, Other)

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 95% CIs and p-value.

- OS analysis using stratification factors as obtained from the baseline CRF pages or database (instead of IRT). This analysis will be performed only if the stratification variable/factor at randomization (as per IRT) and baseline are discordant for at least 10% of randomized subjects. Stratified log-rank test p-value will be generated.
- OS analysis using an un-stratified Cox model. Unstratified log-rank test p-value will be generated.
- OS analysis for participants with no relevant deviation. This analysis will be conducted only if there are more than 10% participants with relevant protocol deviations. Stratified log-rank test p-value will be generated.
- OS in treated subjects from concurrently randomized arms B and C using treatment group "as treated" if more than 10% randomized subjects in any treatment group were never treated or treated differently than randomized among corresponding analysis population. Stratified log-rank test p-value will be generated.
- In order to assess the potential variability introduced by the optional adjuvant chemotherapy:
  - Baseline demographics and disease characteristics, characteristics at the time of surgery will be tabulated by adjuvant therapy status.
  - A Cox regression model with treatment and an additional time-dependant covariate as indicator of the start of adjuvant therapy will used.
  - Additional sensitivity analyses may be performed
- Given the small number of subjects expected to be impacted by a death due to Covid-19 infection, no specifc sensitivity analysis is currently planned. Subjects with death (potentially) associated with Covid-19 infection will be reported based on the reason for death in the death listing.

7.6.4.3 Subset Analyses of OS

The influence of baseline and demographic characteristics on the treatment effect will be explored via subset analyses for the factors specified in section 7.6.2.3.

A forest plot of the OS unstratified hazard ratios (HR) along with two-sided 95% CIs will be produced for each level of the subgroups listed in section 7.6.2.3. If subset category has less than 10 subjects per treatment group, HR will not be computed/displayed. Median and 95% CI will be provided.

In addition, for gender, baseline disease stage, histology, PD-L1 and tumor TMB subsets, Kaplan Meier Curves will be generated.

Kaplan Meier curves, medians and 95% confidence interval will be generated for EFS based on BICR assessments by PD-L1 (≥ 1%, <1%, 1%-49%, ≥ 50%), for Arm A.

7.6.4.4 OS Landmark Analyses by pCR and MPR Status

OS Kaplan-Meier curves will be generated by pCR and by MPR status. These analyses will be landmarked at the time of surgery and will be limited to subjects with pCR or MPR status available. Median and 95% CI will be provided. HR and 95% CI for concurrently randomized subjects in arms B and C will be provided by pCR and by MPR status, as well as HR of pCR vs no pCR by treatment arm.

In addition, OS Kaplan-Meier curves will be generated by pCR and by MPR status, without landmark. These analyses will include all randomized subjects concurrently ranomdized to Arms B and C. Median and 95% CI will be provided. HR and 95% CI for concurrently randomized subjects in arms B and C will be provided by pCR and by MPR status, as well as HR of pCR/MPR vs no pCR/MPR by treatment arm.

If subset category has less than 10 subjects per treatment group, HR will not be computed/displayed.

7.6.4.5 Subject Follow-Up

The extent of follow-up defined as the time between randomization date and last known date alive (for subjects who are alive) or death date (for subjects who died). It will be summarized descriptively (median, min, max, etc) in months.

The currentness of follow-up for survival, defined as the time between last OS contact (i.e., last known date alive or death date) and cut-off date (defined by last patient last visit date), will be summarized in months by treatment group. Subjects who died and subjects with a Last Known Date Alive on or after data cut-off date will have a zero value for currentness of follow-up.

7.6.5 Interim Analyses of EFS and OS

An independent statistician external to BMS will perform the interim analyses. In addition to the formal planned interim analyses for EFS and 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.

Details of interim analyses timing and significance boundaries are provided in sections 5.2 and 5.3.

The DMC will review the safety and available efficacy data as planned in the DMC charter 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.

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

At the time of the interim analysis for of EFS and OS, the DMC may recommend continuing or declare superiority. If the trial continues beyond the formal interim analyses, the nominal critical point for the final analysis will be determined using the recalculated information fraction at the time of the interim analysis, as described above. The final 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 EFS is significant but not OS, the trial is successful but will continue for further OS evalution.

If the p-value crosses the boundary at the interim analysis (EFS or OS), the p-value from the interim stratified log-rank test will be considered the final analysis result.

7.6.6 Analysis of TTDM

Time to death or distant matastases is a secondary endpoint.

TTDM, based on BICR assessments, will be estimated using the Kaplan Meier techniques and will be displayed graphically. A two-sided 95% CI for median TTDM in each treatment group will be computed via the log-log transformation method. TTDM rates at fixed time points (e.g. 6, 12 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

7.6.7 Event Free Survival on Next Line of Therapy

Event Free survival on next line of therapy is an exploratory endpoint.

Event free survival will be estimated using the Kaplan Meier techniques and will be displayed graphically. A two-sided 95% CI for median in each treatment group will be computed via the log-log transformation method. Events rates at fixed time points (e.g. 6, 12 months, depending on the minimum follow-up) will be presented along with their associated 95% CIs. These estimates will be derived from the Kaplan Meier estimate and corresponding CIs will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.

7.6.8 Consistency of Efficacy by Randomization Period

Consistency of population amongs the different randomization periods will be examined by summarizing the primary endpoints by treatment arm as randomized in the the different randomization period: before revised protocol 02 (1:1 A vs B), under revised protocol 02 (1:1:1 A vs B vs C) and after revised protocol 03 (1:1 B vs C). Descriptive statistics will be provided including pCR rates, EFS medians, 95% CI and Kaplan-Meier plots. HR and 95% CI will be provided for arms C vs B during the 1:1:1 randomization (under revised protocol 02) and during the 1:1 randomization (under revised protocol 03).

Concurrent randomization is considered at the site level basis, when the site switched to the revised protcol. In practice, this includes subjects randomized on the randomization lists from the 1:1:1 randomization (revised protocol 02) and the subsequent 1:1 randomization between B and C only (revised protocol 03).

7.7 Safety

Analyses in this section will be tabulated for all treated subjects by treatment group as treated, unless otherwise specified.

Analyses will be performed on the treated subjects from arm A and from the concurrently randomized arms B and C. Participants in Arm B randomized in the initial protocol will only be reported in listings.

7.7.1 Definitive Surgery Related Safety

Incidence of AE/SAE indicated as surgical complication in the CRF, up to 90 days after surgery will be summarized by worst CTC grade, by treatment group.

Adverse events leading to cancellation of surgery and leading to surgery delay will be summarized by worst CTC grade, by treatment group.

7.7.2 Deaths

Deaths will be summarized by treatment group:

- All deaths, reasons for death.
- Deaths within 30 days of last neoadjuvant dose received (30-day safety window), reasons for death.
- Deaths within 100 days of last neoadjuvant dose received (100-day safety window), reasons for death.
- Deaths within 30 days and 90 days from surgery
- Overall summary of AEs leading to death within 100 days of last neoadjuvant dose received
- Overall summary of drug-related AEs leading to death within 100 days of last neoadjuvant dose received
- Overall summary of SAEs leading to death within 100 days of last dose received
- For subjects with systemic adjuvant therapy: deaths from start of systemic adjuvant to last systemic adjuvant dose received + 30 days

A by-subject listing of deaths will be provided for the all enrolled subjects population.

7.7.3 Serious Adverse Events

Serious adverse events will be summarized by treatment group:

- Overall summary of SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- For subjects with systemic adjuvant treatment: Overall summary of SAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT from start of systemic adjuvant to last systemic adjuvant dose received + 30 days
- All analyses will be conducted using the 30-day safety window.

A by-subject SAE listing will be provided for the “enrolled subjects” population.

7.7.4 Adverse Events Leading to Discontinuation of Study Therapy

AEs are indicated as leading to discontinuation, when they lead to discontinuation of at least one agent of the regimen. Reporting is done based on AE CRF form.

AEs leading to discontinuation (of neoadjuvant treatment) will be summarized by treatment group:

- Overall summary of AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of drug-related AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- For subjects with systemic adjuvant treatment: Overall summary of AEs leading to discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT from start of systemic adjuvant to last systemic adjuvant dose received + 30 days

The analyses will be conducted using the 30-day safety window.

A by-subject AEs leading to discontinuation listing will be provided.

7.7.5 Adverse Events Leading to Dose Modification

AEs leading to dose delay/reduction will be summarized by treatment group:

- Overall summary of AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.
- Overall summary of related AEs leading to dose delay/reduction by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

The analysis will be conducted using the 30-day safety window.

A by-subject AEs leading to dose delay/reduction listing will be provided.

7.7.6 Adverse Events

Adverse events will be summarized by treatment group.

The following analyses will be conducted using the 30 days safety window only:

- Overall summary of any AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.
- Overall summary of any AEs presented by worst CTC grade (any grade, grade 3-4, grade 5) by SOC/PT. This table will be restricted to events with an incidence greater or equal to 5% in any treatment group.
- for subjects with systemic adjuvant: Overall summary of any AEs presented by worst CTC grade (any grade, grade 3-4, grade 5) by SOC/PT, from start of systemic adjuvant to last systemic adjuvant dose received + 30 days
- Overall summary of any non-serious AEs presented by SOC/PT. This table will be restricted to events with an incidence greater or equal to 5% in any treatment group.
- Overall summary of any AEs that required immune modulating medication by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT.

- Overall summary of drug-related AEs by worst CTC grade (1, 2, 3, 4, 5, not reported, total) presented by SOC/PT.
- For subjects with systemic adjuvant treatment: Overall summary of Drug-Related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT from start of systemic adjuvant to last systemic adjuvant dose received + 30 days

The following analyses will be conducted using the 30 days safety window and repeated using the 100 days safety window:

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

A by-subject AE listing will be provided. A by-subject listing of any AE requiring immune modulating medications will also be provided.

7.7.7 Select Adverse Events

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

7.7.7.1 Incidence of Select AE

Select AEs will be summarized by treatment group for each category/subcategory.

The following analyses will be conducted using the 30-day safety window only:

- 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 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 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-subject select AE listing will be provided.

7.7.7.2 Time-to Onset of Select AE

Time-to onset of drug-related select AEs (any grade, grade 3-5) will be summarized for each category/subcategory by treatment group.

Time-to onset analyses are restricted to treated subjects who experienced at least one drug-related select AE in the category/subcategory. The analyses will be conducted using the 30-day safety window.

Additional details regarding the time-to onset definition are described in time-to onset definition subsection of APPENDIX 1.

7.7.7.3 Time-to Resolution of Select AE

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

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

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

The analyses will be conducted using the 30-day safety window.

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

See time-to resolution definition subsection of APPENDIX 1 for additional details.

7.7.8 Immune-Mediated Adverse Events

IMAEs will be summarized by treatment group for each immune-mediated category / PT using the 100-day safety window:

- 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 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 discontinuation by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT.
- Overall summary of non-endocrine IMAEs leading to 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 endocrine IMAEs leading to dose delay or 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-subject listing of IMAEs will be provided. By-subject listings of time-to resolution for longest IMAEs cluster (any grade and grade 3-5 in separate summaries) will also be provided. For new studies which collect investigator assessment of potential IMAE data, a by-subject listing of AEs considered as immune-mediated events per investigator but not qualified for IMAEs definition will also be provided.

7.7.9 Other Events of Special Interest

OEOSI will be summarized by treatment group for each category.

The following analyses will be conducted using the 100-day safety window:

- 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-subject listing of OEOSI will be provided.

7.7.10 Multiple Events

The following summary tables will be provided:

- A table showing the total number and rate (exposure adjusted) of occurrences for all AEs.
- A table showing the total number and rate (exposure adjusted) of occurrences for AEs occurring in at least 5% of subjects in any treatment group.

Exposure adjustement will be based on the exposure in neoadjuvant treatment only.

A listing displaying the unique instances of all AEs, i.e., after duplicates have been eliminated and overlapping and contiguous occurrences of the same event (i.e. same PT) have been collapsed will be provided. No formal comparisons will be made between treatment groups.

7.7.11 New Primary Cancers

Occurence of new primary cancer reported on the new primary cancer CRF form will be listed based on the all treated population.

7.7.12 Laboratory Parameters

The analysis population for each laboratory test is restricted to treated subjects who underwent that laboratory test.

A by-subject listing of differences in categorization of SI and US laboratory test results will be provided.

7.7.12.1 Hematology

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: hemoglobin (HB), platelets, white blood counts (WBC), absolute neutrophils count (ANC) and lymphocyte count (LYMPH).

The analyses will be conducted using the 30-day safety window.

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

7.7.12.2 Serum Chemistry

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: ALT, AST, alkaline phosphatase (ALP), total bilirubin, creatinine, amylase, lipase.

The analyses will be conducted using the 30-day safety window.

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

7.7.12.3 Electrolytes

The following will be summarized by treatment group as worst CTC grade on-treatment per subject and as shift table of worst on-treatment CTC grade compared to baseline CTC grade per subject: sodium (high and low), potassium (high and low), calcium (high and low), magnesium (high and low), and Glucose Serum (fasting hyperglycemia and hypoglycemia regardless of fasting status).

The analyses will be conducted using the 30-day safety window.

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

7.7.12.4 Additional Analyses

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

Abnormal Hepatic Function Test

The number of subjects 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
- ALP > 1.5 x ULN
- Concurrent (within 1 day) 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 > 1.5 x ULN
- Concurrent (within 1 day) 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

The analyses will be conducted using the 30-day safety window.

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

Abnormal Thyroid Function Test

The number of subjects 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 30-day safety window.

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

7.7.13 Vital Signs and Pulse Oximetry

Vital signs and pulse oximetry (i.e. % oxygen saturation) collected on the CRF will be provided in separate listings.

7.7.14 Physical Measurements

Physical measurements will be listed by subject.

7.7.15 Non-Protocol Medical Procedures

Non-protocol medical procedures will be listed by subject.

7.7.16 Immunogenicity Analysis

Not applicable in this protocol.

7.7.17 Pregnancy

A by-subject listing of pregnancy tests results will be provided for randomized female subjects.

7.7.18 Adverse Events By Subgroup

Overall summary of any AEs and drug-related AEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by SOC/PT and for each treatment group for the following subgroups:

- Sex (Male vs. Female)
- Race
- Age (< 65 vs. 65 - < 75 vs. 75 - < 85 vs. ≥ 85 vs. ≥ 75 vs. ≥ 65)
- Region (North America, Europe, Asia, Rest of World)
- Type of platinum therapy (cisplatin, carboplatin, subjects switching from cisplatin to carboplatin).

- Type of chemotherapy regimen in arm B (available in arm C (Gemcitabine-Cisplatin, Pemetrexed-Cisplatin, Paclitaxel-Carboplatin, not available in arm C (Vinorelbine-Cisplatin, Docetaxel-Cisplatin)) , based on first neoadjuvant cycle.

These analyses will be conducted using the 30-day safety window only.

7.7.19 Consistency of Safety by Randomization Period

Consistency of population amongs the different randomization periods will be examined by summarizing key safety by treatment arm as randomized in the the different randomization period: before revised protocol 02 (1:1 A vs B), under revised protocol 02 (1:1:1 A vs B vs C) and after revised protocol 03 (1:1 B vs C).

- Overall summary of any AEs presented by worst CTC grade (any grade, grade 3-4, grade 5) by SOC/PT with 30-day safety window.
- Overall summary of AEs leading to discontinuation (of neoadjuvant treatment) by worst CTC grade(any grade, grade 3-4, grade 5) presented by SOC/PT with 30-day safety window.
- 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 with 100-day safety window.
- Overall summary of endocrine IMAEs by worst CTC grade (any grade, grade 3-4, grade 5) presented by Category / PT with 100-day safety window.
- Concurrent randomization is considered at the site level basis, when the site switched to the revised protcol. In practice, this includes subjects randomized on the randomization lists from the 1:1:1 randomization (revised protocol 02) and the subsequent 1:1 randomization between B and C only (revised protocol 03).

7.7.20 Covid-19

- COVID-19 related adverse events will be summarized (100 days window) and listed.

7.8 Pharmacokinetics

The nivolumab and ipilimumab concentration data obtained in this study will be combined with data from other studies in the clinical development program to develop a population PK model. This model will be used to evaluate the effects of intrinsic and extrinsic covariates on the PK of nivolumab and ipilimumab. In addition, exposure-response analyses with selected efficacy and safety endpoints will be conducted. Results of population PK and exposure response-analyses will be reported separately.

7.9 Biomarkers

Analyses for PD-L1 and TMB (tumor tissue) are described below.

Methodology for biomarkers other than PD-L1,Tumor TMB [redacted] will be reported separately.

7.9.1 PD-L1 Expression

Descriptive statistics of PD-L1 expression:

- Listing of all PD-L1 IHC data, all randomized subjects.

- Summary of tumor specimen acquisition and characteristics, all randomized subjects.
- Summary statistics of PD-L1 expression in all randomized subjects with quantifiable PD-L1 expression.
- Cumulative distribution plot of baseline PD-L1 expression versus population percentile in all randomized subjects with quantifiable PD-L1 expression.
- Box plots of PD-L1 expression versus pCR in all randomized subjects with quantifiable PD-L1 expression.
- Waterfall plot of Individual PD-L1 expression in all randomized subjects with quantifiable PD-L1 expression.
- Subgroups analyses of efficacy (pCR, MPR, EFS, OS) by PD-L1 status (PD-L1<1% PD-L1 ≥ 1%, PD-L1 1-49%, PD-L1 ≥ 50%, not evaluable/indeterminate) are described in Section 7.6.
- In addition the following analyses will be conducted in the concurrently randomized subjects in arms B and C.
- An exploratory Cox proportional hazards model, in order to assess the association of EFS (based on BICR assessments) with PD-L1, will be fitted for EFS with PD-L1, treatment arm and PD-L1* treatment arm interaction, among all PD-L1 evaluable subjects. An appropriate transformation of PD-L1 expression may be considered depending on an assessment of fit of the model. The following summaries will be presented by treatment arm
  - A plot of estimated $\log_e$(hazard ratio) with 95% confidence band vs PD-L1 expression(X-axis)
- An exploratory Cox proportional hazards model, in order to assess the association of OS with PD-L1, will be fitted for OS with PD-L1, treatment arm and PD-L1* treatment arm interaction, among all PD-L1 evaluable subjects. An appropriate transformation of PD-L1 expression may be considered depending on an assessment of fit of the model. The following summaries will be presented by treatment arm
  - A plot of estimated $\log_e$(hazard ratio) with 95% confidence band vs PD-L1 expression (X-axis).
- A logistic regression model will be fitted for pCR with PD-L1, treatment arm and PD-L1* treatment arm interaction, among all PD-L1 evaluable subjects. The following summaries will be reported:
  - A plot of estimated odds ratio with 95% confidence band vs PD-L1 expression (X-axis)
- For both EFS (using primary definition, per BICR) and OS, a Cox proportional hazards regression model will be fitted with treatment, PD-L1 status (using 1%, 5%, 50% cut offs), and treatment by PD-L1 status interaction. Although the study is not designed to have appropriate power to formally test the interaction of the model, an interaction test at significance level of 0.2 will warrant further exploration and the following statistics will be reported:
  - Interaction p-value
  - HR of treatment vs. control and its associated 95% CI for each of the PD-L1 status subgroup
  - HR PD-L1 ≥ cutoff vs. < cutoff and its associated 95% CI within each treatment group

7.9.2 Tumor Mutational Burden (TMB)

The analyses are based on tumor tissue TMB evaluable subjects, defined as subjects with tissue TMB data available. It is known that not all subjects will provide tumor tissue TMB data, due to factors such as available remaining tissue and inherent failure rates of the TMB process.

The descriptive analyses of tumor tissue TMB at baseline will be conducted:

- Listing of all tumor tissue TMB data.
- Summary of tumor specimen characteristics
- Cumulative distribution plot of TMB at baseline versus population percentile in all subjects with evaluable tumor tissue TMB.

In addition, the joint distribution of PD-L1 and tumor tissue TMB among both PD-L1 and tumor tissue TMB-evaluable subjects will be examined.

Subgroups analyses of efficacy (pCR, MPR, EFS, OS) by TMB status (≥12.3 mut/MB, < 12.3 mut/MB, TMB evaluable, TMB not evaluable) are described in Section 7.6.

- In addition the following analyses will be conducted in the concurrently randomized subjects in arms B and C.
- An exploratory Cox proportional hazards model, in order to assess the association of EFS (based on BICR assessments) with TMB, will be fitted for EFS with TMB, treatment arm and TMB* treatment arm interaction, among all TMB evaluable subjects. An appropriate transformation of TMB may be considered depending on an assessment of fit of the model. The following summaries will be presented by treatment arm
  - A plot of estimated $\log_e$(hazard ratio) with 95% confidence band vs PD-L1 expression(X-axis)
- An exploratory Cox proportional hazards model, in order to assess the association of OS with TMB, will be fitted for OS with TMB, treatment arm and TMB* treatment arm interaction, among all TMB evaluable subjects. An appropriate transformation of TMB may be considered depending on an assessment of fit of the model. The following summaries will be presented by treatment arm
  - A plot of estimated $\log_e$(hazard ratio) with 95% confidence band vs TMB (X-axis).
- A logistic regression model will be fitted for pCR with TMB, treatment arm and TMB* treatment arm interaction, among all TMB evaluable subjects. The following summaries will be reported:
  - A plot of estimated odds ratio with 95% confidence band vs TMB (X-axis)

[redacted]

[redacted]

[redacted]

7.10 Outcomes Research Analyses

The analysis of EQ-5D-3L will be restricted to randomized subjects who have an assessment at baseline and at least one post-baseline assessment.

The following descriptive analyses will be conducted:

- EQ-5D-3L questionnaire completion rate, defined as the proportion of questionnaires actually received out of the expected number (i.e. number of subjects on treatment or in follow up), will be calculated and summarized for each assessment time point by treatment group.
- A by-subject listing of the level of problems in each dimension, corresponding to EQ-5D-3L health state (i.e., 5-digit vector), EQ-5D-3L utility index score, and EQ-5D-3L VAS score will be provided.
- Proportion of subjects reporting problems for the 5 EQ-5D-3L dimensions at each assessment time point will be summarized by level of problem and by treatment group. Percentages will be based on number of subjects assessed at assessment time point.
- For the EQ-5D-3L utility index and VAS scores, separately:
  - Mean score and mean change from baseline at each assessment time point will be summarized by treatment group using descriptive statistics (N, mean with SD and 95% CI, median, first and third quartiles, minimum, maximum).
  - A line graph summarizing the mean changes from baseline will be produced.

7.11 Country Specific Analyses

Country or region specific analyses may take place to support regional health authorities submissions. In general these would consist of repeating a subset of the analyses described in this SAP for a subjects from a specific country or region, which would potentially be used for evaluation of consistency between the country/region and the overall population. Unless otherwise noted, no formal hypothesis testing will be performed to evaluate consistency of the subpopulation. Instead, descriptive statistics will be provided to assess the consistency.

These analyses provided for regional submission needs would only performed when at least one of the primary endpoints is statistically significant (based on data from the global population), for regional submission needs. The analysis methods for the subpopulation will be the same as for the global population unless otherwise noted.

For China, the the following analysis sets will be used:

China subpopulation

Chinese subpopulation is defined as Great China subgroup. It should contain subjects that are Chinese by race and enrolled from China, Hong-Kong or Taiwan sites.

Asian subpopulation

Asian subpopulation is defined as Asian subgroup. It should contain subjects that are Asian by race and enrolled from Asian countries, as per this SAP geographic region derivation.

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 Specification17
- For missing and partial adverse event resolution dates, imputation will be performed as follows:
  - If only the day of the month is missing, the last day of the month will be used to replace the missing day. If the imputed date is after the death date or the last known alive date, then the latest known alive date or death date is considered as the resolution date.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- Missing and partial non-study medication domain dates will be imputed using the derivation algorithm described in 4.1.3 of BMS Non-Study Medication Domain Requirements Specification[18].
- Missing and partial radiotherapy and surgery dates will be imputed using algorithm described in APPENDIX 2.
- Missing of partial definitive surgery date
  - If only the day of the month is missing, the 1st of the month will be used to replace the missing day. In case of the date of death is present and complete, the imputed definitive surgery date will be compared to the date of death. The minimum of the imputed definitive surgery date and date of death will be considered as the date of definitive surgery.
  - If the day and month are missing or a date is completely missing, it will be considered as missing.
- 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 alive date 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 alive date.
  - 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/recurrence after start of study therapy, 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. In case of the date of death is present and complete, the imputed progression/recurrence date will be compared to the date of death. The minimum of the imputed progression/recurrence date and date of death will be considered as the date of progression/recurrence.

  - If the day and month are missing or a date is completely missing, it will be considered as missing.

- For date of progression to prior therapies, 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.
- For other partial/missing dates, the following conventions were 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-to onset, time-to resolution) will be calculated as follows:

Duration = (Last date - first date + 1)

Last known alive date will be defined based on all appropriate dates collected on the CRF.

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. Additional exploratory analyses may be performed. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

10 DOCUMENT HISTORY

Table 10-1: Document History

Version Number	Author(s)	Description
1.0	[redacted]	Initial version dated 17-Apr-2020
2.0	[redacted]	29-Jul-2020 Revised Protocol 06 related changes: • Clarified EFS definition

Table 10-1: Document History

Version Number	Author(s)	Description
		• Removed the 1st interim analysis of EFS and updated alpha spending of interim and final analyses of EFS, based on health authorities feedback • Clarified that actual timing of analyses may differ from projected timing <u>Other changes:</u> • Removed immunogenicity analyses as immunogenicity is not described in the protocol. • Sections 4.1.2 and 4.2.2, clarified that subjects with pCR/MPR not evaluable/no avaialble are considered as non-responders • Section 4.3.3: specified that the length of surgery delay will be reported [redacted] • Section 4.3.7: EQ-5D analyses: clarify that the windowing for adjuvant is based on systemic adjuvant only • Section 5.3: removed OS analyses methodology which is repeated in section 7 • Section 5.4: Refer to the DMC charter for handling of EFS descriptive analysis at the time of pCR • Section 7.2.2: Relevant deviations: clarify that the concurrent therapies are limited to neoadjuvant and adjuvant systemic treatment • Section 7.3.2: corrected the PDL1 should be per clinical database • Section 7.3.4: removed reporting of prior systemic therapies, radiotherapy and surgeries since the study is in neoadjuvant setting • Section 7.4.2.1: Dose delay/omission: describe the derivation of delay, when there is an omission between two doses. and the fact that delays during the adjuvant treatment are limited to reported delays on the CRF. • Section 7.5: added duration of surgery delay • Section 7.6: Added Covid-19 related consideration in the efficacy supportive analysis sections • Section 7.6.2.2: added pCR sensitivity analysis using CRF stratification factors • Section 7.6.3.2: added sensitivity analysis considering progressions before surgery as events, even if not precluding surgery • Section 7.6.3.4 and 7.6.4.4: added calculation of HR for pCR vs no pCR. • Section 7.6.4.2: remove extra paragraph on analysis on concurrently randomized arms B and C. • Section 7.6.5: clarified that the DMC may declare superiority but not stop the trial for superiority as the trial will continue for further endpoints • Section 7.7: clarified that the window for reporting AEs during the adjuvant period is based on systemic aduvant theratpy (not radiotherapy) • Added section 7.11 for country/region specific analyses

Table 10-1: Document History

Version Number	Author(s)	Description
3.0	[redacted]	• The main reason for amendment is to reflect the changes to Revised Proocol 07: add an interim analysis for EFS/OS [redacted] The rationale for change is the potential for a slow down of EFS events with a plateau towards the end of the curve. • [redacted] as well as some additional sensitivity, subgroup analyses and clarifications. • Detailed changes are provided below: • Section 1: removed reference to the Core SAP which is outdated • Section 1: updated the schedule of analyses with an additional EFS/OS interim analysis at 90% EFS information fraction [redacted] • Section 2.3: updates: In case of non-significant EFS at interim analysis, the DMC unblinded closed report will be shared with BMS executive restricted team. Further details on data communication are provided in section 2.3 and DMC charter. • Section 2.4: updated the Revised Protocol 07 and added rationale for revised protocol 05 (in response to internal audit finding) [redacted] • Section 5.2: Updated with revised EFS assumptions assuming piecewise exponential distribution and added EFS interim analysis [redacted] • Section 5.3: updated to add one OS interim analysis at the time of newly added EFS interim analysis • Section 5.4: updated analyses timing following changes in section 5.2 and indicated that the the DMC closed report, including OS will be shared with the BMS restricted team. removed table 5.4-1 which was repeating information from Table 5.3-1 and section 5.4 • Section 7.3.1: clarified that covid-19 related reason will be provided • Section 7.4.2: clarified that covid-19 related reason will be provided • Section 7.6.3.2: Clarified that for EFS per investigator the censoring for new primary cancer is at the time of last tumor assessment prior or at the time of new primary cancer, instead of the date of new primary cancer • Section 7.6.3.2: added sensitivity analysis with interval censoring method • Section 7.6.3.3: added EFS analysis by PD-L1 for Arm A • Section 7.6.3.4: clarified that no HR will be generated for subsets with less than 10 subjects, added HR by MPR • Section 7.6.3.4: added EFS analysis by pCR and MPR without landmark to inform surrogacy

Table 10-1: Document History

Version Number	Author(s)	Description
		• Section 7.6.4.3: added OS analysis by PD-L1 for Arm A • Section 7.6.4.4: clarified that no HR will be generated for subsets with less than 10 subjects, added HR by MPR • Section 7.6.4.4: added OS analysis by pCR and MPR without landmark to inform surrogacy • Section 7.7.2: added Death within 30 days and 90 days from surgery and added AE/SAE/related AE leading to death • Section 7.7.20: added section on Covid-19 AEs [redacted] • Section 7.9.1: added predictive analysis between EFS/OS and PD-L1 [redacted] • Section 11: updated the list of prior analysis including the pCR analysis

11 PREVIOUS ANALYSES

The following DMC safety meetings occured before approval of this SAP. Analyses were generated for these meetings according to the DMC charter.

- 1st interim safety review for Arms A and B after approximately 15 subjects in each arm completed surgery (14-Mar-2018)
- 2nd interim safety review after approximately 15 subjects in arm C completed surgery (03-Oct-2018)
- Additional safety reviews approximately every 6 months until the primary endpoint of pathological complete response is analyzed (24-May-2019, 04-Dec-2019)

No by treatment data have been share with BMS for these meetings.

The analysis of the pCR primary endpoint was performed with a database lock of 16-Sep-2020. The DMC reviewed the data and BMS was unblinded to treatment arm. A Clinical Study Report was written reporting the demographics, baseline characteristics, surgery related data, pathological response results, clinical response and safety. No EFS/OS data by treatment arm was shared with BMS.

APPENDIX 1 TIME-TO ONSET AND TIME-TO RESOLUTION DEFINITION AND CONVENTIONS FOR SELECT ADVERSE EVENTS, IMMUNE-MEDIATED ADVERSE EVENTS AND EVENTS OF SPECIAL INTEREST

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.

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 subject (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 1 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 subject 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 subjects 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 subject 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 1: 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 subject and specified category, the corresponding adverse event records will be collapsed when:

3) Multiple adverse event records have the same onset date.
4) 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).
5) 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).

APPENDIX 2 MISSING AND PARTIAL RADIOTHERAPY AND SURGERY DATES IMPUTATION ALGORITHMS

Procedures – Imputation Rules.

If reported procedure start date is a full valid date then set start date equal to the date part of procedure start date.

In case of partial date use imputation rules described below:

- If only day is missing then
  - If month and year of procedure match month and year of first dose date then impute as date of first dose;
  - If month and year of procedure don't match month and year of first dose date then impute as first day of that month and year.
- If both day and month are missing, then impute as maximum between 01JAN of the year and date of the first dose;
- If date is completely missing or invalid then leave missing.

Note: Imputation is not applicable to data where start date is not collected (for example "PRIOR RADIOTHERAPY" CRF). Set start date to missing in this case.

If reported end date is a full valid date then set end date equal to the date part of the reported end date.

In case of partial date use imputation rules described below:

- If reported end date is partial then set end date equal to the last possible reported end date based on the partial entered reported end date.
- If reported end date is missing, continuing, unknown or invalid then set end date equal to the most recent database extraction date.

If end date was imputed then compare end date to the death date or last known alive date if subject is not dead. If posterior then end date should be imputed to death date (or last known alive date if subject not dead).

Note: Imputation of partial dates only applies to data entered on "RADIOTHERAPY" CRF page. For other CRF pages in case of partial dates set end date to missing.

Surgeries – Imputation Rules.

If reported surgery date is a full valid date then set start date equal to the date part of surgery date.

In case of partial date, use one of the two imputation rules described below:

A. For data collected on "PRIOR SURGERY RELATED TO CANCER" CRF page:

- If only day is missing then impute as the first day of the month;
- If both day and month are missing then then impute as 01JAN of the year;
- If date is completely missing or invalid then leave missing.

B. For data collected on “SUBSEQUENT SURGERY” CRF page:

- If only day is missing then
  - If month and year of surgery match month and year of first dose date then impute the missing date as the date of first dose;
  - If month and year of surgery don't match month and year of first dose date then impute as first day of that month and year;
- If both day and month are missing then impute as maximum between 01JAN of the year and date of the first dose;
- If date is completely missing or invalid then leave missing.
-
- C. For DEFINITIVE SURGERY CRF page :
- If only day is missing then
  - if month and year of surgery match month and year of first dose date then impute the missing date as the date of first dose;
  - if month and year of surgery match month and year of last neoadjuvant dose date then impute the missing date as the date of last Neoadjuvant dose
  - if month and year of surgery don't match month and year of first dose or last Neoadjuvant dose date then impute as first day of that month and year;
- If both day and month are missing then impute as maximum between 01JAN of the year and date of last Neoadjuvant dose date;
- If date is completely missing or invalid then leave missing.
- For incomplete definitive surgery end date: set to definitive surgery start date

12 REFERENCES

1 Mountain, CA. Regional Lymph node classification for lung cancer staging. Chest. 1997: 1718-23.

2 Goldstraw P, Crowley J, Chansky K, et al for the International Association for the Study of Lung Cancer International Staging Committee. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage grouping in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Onco. 2007; 2(8): 706-14

3 Detterbeck FC, Boffa DJ, Tanoue LT. The new lung cancer staging system. Chest. 2009; 136: 260-71.

4 Provencio et al. Abs 8521 ASCO 2018

5 Baden et al, Comparison of Platforms for Determining Tumor Mutational Burden in Patients With Non-Small Cell Lung Cancer, ESMO 2019

6 Dolan P. Modeling valuations for EuroQol health states. Medical Care 1997;35: 1095-1108

7 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.

8 Glimm E., Maurer W., Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat. Med. 2010;29:219–228.

9 Clopper, CJ and Pearson, ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26: 404-423, 1934.

10 Brookmeyer R. and Crowley J. A confidence interval for the median survival time. Biometrics 38:29-41, 1982

11 Klein, J. P. and Moeschberger, M. L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag

12 Kalbfleisch RL. The Statistical Analysis of Failure Time Data. Wiley-Interscience, 2002.

13 Global Biometric Sciences, SAS Analysis Dataset Specification, Unique Adverse Events. Version 2.0, April 22, 2015

14 Statistical Methodology in the Pharmaceutical Sciences/ edited by Berry DA, Chapter 13, Categorical Data Analysis, p 415 and 417 ff., Marcel Dekker, 1990.

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

16 Kalbfleisch, J. D. and Prentice, R. L. (1980), The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons.

17 Adverse Event Domain Requirements Specification Bristol Myers Squibb Co. PRI. Version 2.3.0 April 23, 2018

[18] Non-Study Medication Domain Requirements Specification Bristol Myers Squibb Co. PRI. Version 2.10.0 April 23, 2018