Official Title of Study:

A RANDOMIZED, OPEN-LABEL, PHASE 3 TRIAL OF NIVOLUMAB PLUS BRENTUXIMAB VEDOTIN VERSUS BRENTUXIMAB VEDOTIN ALONE IN PARTICIPANTS WITH RELAPSED REFRACTORY OR INELIGIBLE FOR AUTOLOGOUS STEM CELL TRANSPLANT (ASCT) ADVANCED STAGE CLASSICAL HODGKIN LYMPHOMA

PROTOCOL(S) CA209-812

NCT Number: NCT03138499

Document Date (Date in which document was last revised): February 24, 2017

**STATISTICAL ANALYSIS PLAN
FOR CLINICAL STUDY REPORT**

*A RANDOMIZED, OPEN-LABEL, PHASE 3 TRIAL OF NIVOLUMAB PLUS
BRENTUXIMAB VEDOTIN VERSUS BRENTUXIMAB VEDOTIN ALONE IN
PARTICIPANTS WITH RELAPSED REFRACTORY OR INELIGIBLE FOR
AUTOLOGOUS STEM CELL TRANSPLANT (ASCT) ADVANCED STAGE
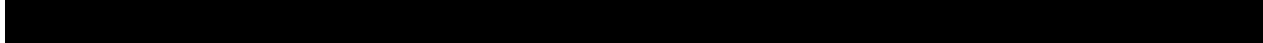CLASSICAL HODGKIN LYMPHOMA*

*CHECKMATE 812: CHECKPOINT PATHWAY AND NIVOLUMAB CLINICAL TRIAL
EVALUATION 812*

PROTOCOL CA209-812

VERSION # 1.0

TABLE OF CONTENTS

Approved v 3.0

LIST OF TABLES

LIST OF FIGURES

1   BACKGROUND AND RATIONALE

CA209-812 is a randomized, open-label, Phase 3 Trial of nivolumab plus brentuximab vedotin (nivolumab+BV) versus brentuximab vedotin (BV) in participants who are relapsed/refractory or ineligible for autologous stem cell transplant (ASCT). The central question of the study will be to determine if nivolumab+BV improves progression free survival (PFS) over BV in this patient population. Additional objectives include further characterization of the efficacy ███████████ ██ of nivolumab+BV vs. BV in this population.

This statistical analysis plan (SAP) details all analyses that are planned in the Clinical Study Report (CSR) for CA209812 study except for safety which will be documented separately in the Core Safety SAP[1].

Research Hypothesis:

Nivoluamb+BV will provide benefit over BV, as evidenced by improvement in PFS assessed by Blinded Independent Central Review (BICR), to participants who are relapsed/refractory or ineligible for autologous stem cell transplant (ASCT).

Schedule of Analyses:

The interim PFS analysis is scheduled when at least 70% (131/187) of final PFS events have been observed with a minimum follow-up of 6 months since the last patient first visit (LPFV), which is projected to occur approximately 40 months after the start of randomization. The final PFS analysis is scheduled when 187 PFS events have been observed, which would occur approximately 45 months after start of randomization (34 months accrual + 11 months follow up). Additional survival follow-up may continue for up to 4 years from the time of the first dose of the last enrolled participant. The study will end once survival follow-up has concluded.

All secondary endpoints will be analyzed at the time of the interim and final PFS analysis.

2   STUDY DESCRIPTION

2.1   Study Design

This is a 1:1 randomized, open-label Phase 3 study in advanced cHL participants ≥ 18 years old who are relapsed refractory or ineligible for ASCT. Patients will be balanced in the 2 groups in regards to prior therapies. Approximately 340 participants will be treated in one of 2 arms. In one arm, participants will receive nivolumab 360 mg IV every 3 weeks until progression or unacceptable toxicity (except for patients in CR who can discontinue at 2 years) plus BV 1.8 mg/kg IV every 3 weeks for up to 16 cycles, or until disease progression or unacceptable toxicity, whichever occurs first. In the other arm, participants will receive BV alone 1.8 mg/kg every 3 weeks for up to 16 cycles, disease progression, or unacceptable toxicity, whichever occurrs first. Discontinuation of study therapy can also occur if the participant meets other criteria outlined in Protocol Section 8.1. Participants receiving nivolumab who achieve CR may discontinue treatment

after a maximum of 2 years of therapy, provided that there is no prohibitive toxicity. Participants can be brentuximab vedotin-naïve, or can have prior BV treatment as a single agent or in combination in any line of therapy. Randomization stratification will be performed on the following two factors: 1) Prior ASCT status (YES/NO). 2) Prior BV use (YES/NO).

- Duration of study participation: Randomization is within 28 days of the screening. The recruitment period is approximately 34 months with a follow up of 4 years after the last patient first visit (LPFV). Participants will receive active treatment of nivolumab until progression or unacceptable toxicity plus BV for 16 cycles or until progression or unacceptable toxicity, whichever comes first, or BV alone up to 16 cycles, or until progression or unacceptable toxicity, whichever occurs first. Each 21-day dosing period will constitute a cycle. The treatment period will be followed by a survival follow up period.

This study will consist of 3 phases: screening, treatment, and follow-up.

Figure 2.1-1:   **Study Design**
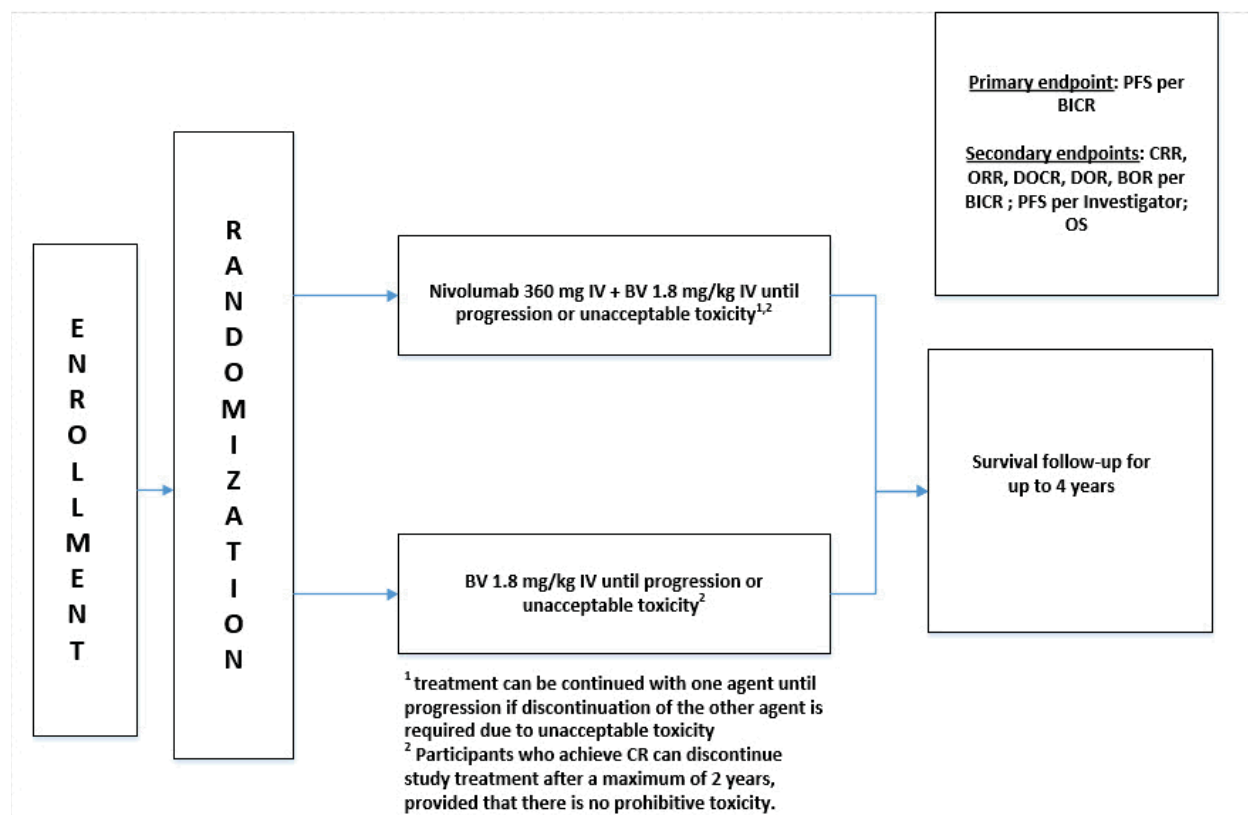
Abbreviations: BV = brentuximab vedotin; BICR = Blinded Independent Central Review; PFS = progression-free survival; CRR = complete response rate; OS = overall survival; ORR = objective response rate; DOCR = duration of complete response; DOR = duration of response; BOR = best overall response

7

2.2   Treatment Assignment

Participants will be identified and informed consent obtained. Participants must be enrolled into the study by the interactive web response system (IVRS) to obtain the participant number. For both randomized arms, the investigator or designee will register the participant for enrollment by following the procedures established by BMS. The following information is required for registration:

- Date that informed consent was obtained
- Date of birth
- Gender at birth

Once registered, participants who have signed the ICF and met all eligibility criteria will be ready to be randomized. The site will enter into the IVRS to obtain the treatment assignment. Participants will be randomly assigned through IVRS in 1:1 ratio to either arm A (nivolumab+BV) or arm B (BV) stratified by two factors: 1) Prior ASCT status (YES/NO), 2) Prior BV use (YES/NO). The randomization will be carried out via permuted blocks within each stratum. The exact procedures for using the IVRS will be detailed in a separate document.

2.3   Blinding and Unblinding

Not applicable. This is an open-label study.

2.4   Protocol Amendments

None.

2.5   Data Monitoring Committee

An independent Data Monitoring Committee (DMC) will be established to provide oversight of safety and efficacy considerations and to provide advice to the Sponsor regarding actions the committee deems necessary for the continuing protection of participants enrolled in the trial. Details of DMC responsibilities and procedures will be specified in the DMC charter.

3   OBJECTIVES

3.1   Primary Objective

- To compare progression free survival of nivolumab+BV vs. BV based on BICR assessments

3.2   Secondary Objectives

Secondary objectives include the following:
- To compare the complete response rate of nivolumab+BV vs. BV based on BICR assessments
- To assess objective response rate and duration of response based on BICR
- To assess duration of complete response based on BICR
- To assess overall survival of participants treated with nivolumab + BV vs BV
- To assess PFS based on investigator assessments

- To assess the overall safety and tolerability of nivolumab in combination with chemotherapy, as measured by incidence and severity of adverse events (AEs), serious adverse events (SAEs), and specific laboratory abnormalities
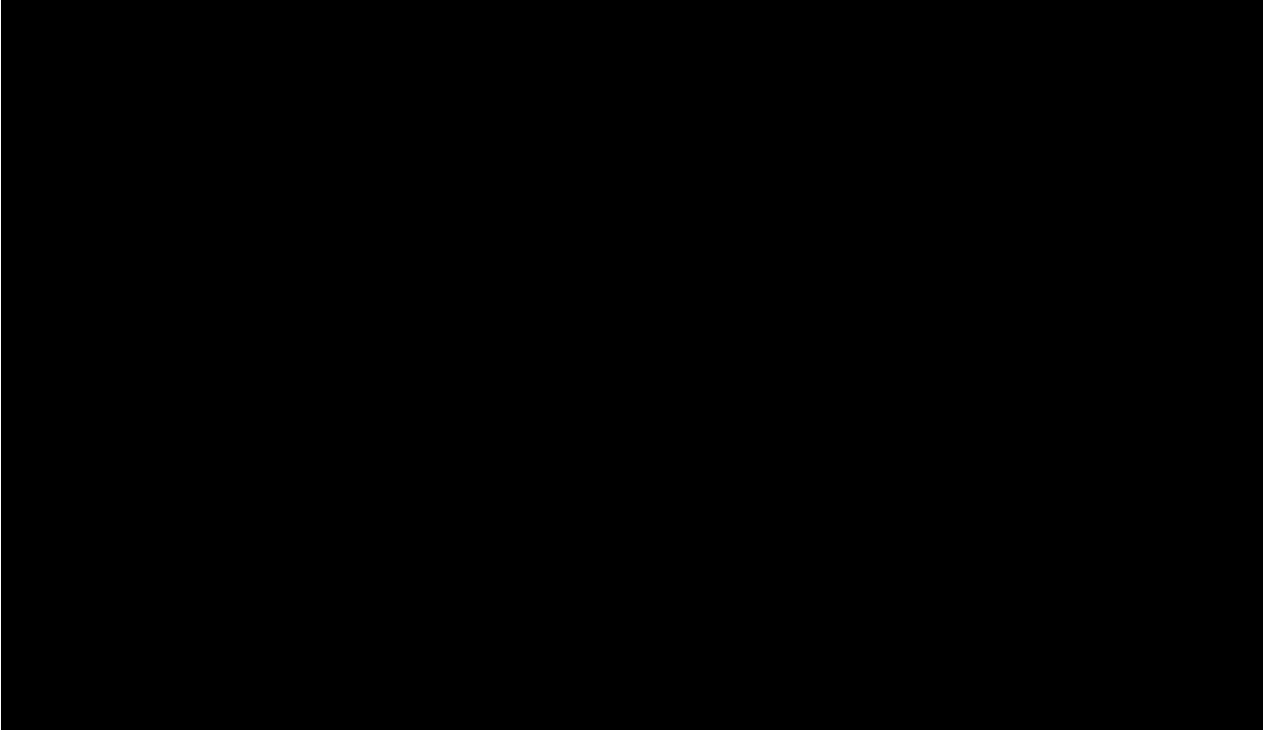
4 ENDPOINTS

4.1 Primary Endpoints

Progression Free Survival (PFS)

Progression-free Survival (PFS) by BICR: defined as time from date of randomization to the date of death due to any cause or the first documented disease progression as determined by the BICR, whichever occurs first.

Participants who die without a reported progression will be considered to have progressed on the date of their death. Participants who did not progress or die will be censored on the date of their last tumor assessment. Participants who did not have any on study assessments and did not die will be censored on the randomization date. A participant who initiates a subsequent anti-cancer therapy (including allogenic transplant) will be censored on the date of last tumor assessment prior or on the start date of their subsequent anti-cancer therapy.

The censoring scheme is summarized in Table 4.1-1.

Table 4.1-1: **Censoring scheme used in primary analysis of PFS**

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization date	Censored

No on-study tumor assessments and no death	Randomization date	Censored
New subsequent anticancer therapy started without a prior reported progression by BICR or death	Date of last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored
Progression per BICR documented at a scheduled or unscheduled visit and no other subsequent anticancer therapy started before	Date of the first documented tumor progression	Progressed
Participant progression free, no death and no subsequent anti-cancer therapy started	Date of last evaluable tumor assessment	Censored
Death without prior progression per BICR and no subsequent anticancer therapy started	Date of death	Progressed

4.2 Secondary Endpoints

4.2.1 BOR, CRR, ORR, PR, TTR, DOR and DOCR

Best Overall Response (BOR) of CR or PR or SD or PD will be evaluated for response evaluable participants according to the 2014 Lugano classification, based on BICR assessment.

The BOR is defined as the best response designation recorded between the date of randomization and the date of initial objectively documented progression per the 2014 Lugano classification or the date of subsequent therapy, whichever occurs first. For participants without documented progression or subsequent therapy, all available response designations will contribute to the BOR determination. For participants who continue treatment beyond progression, the BOR should be determined based on response designations recorded up to the time of the initial BICR defined progression.

Complete Response Rate (CRR): defined as proportion of participants who have achieved complete response (Lugano 2014 conference) as determined by the BICR prior to initiating any subsequent therapy among the randomized participants. For participants who continue treatment beyond progression, the complete response should be determined based on response designations recorded up to the time of the initial BICR defined progression.

ORR is defined as the number of participants with a Best Overall Response (BOR) of CR or PR, per the 2014 Lugano classification, based on BICR assessment, divided by the number of randomized participants

The objective response will be further characterized by the time to response (TTR). TTR is defined as the time from first dosing date to the date of the first response, based on BICR assessment. . TTR will be evaluated for responders only.

DOR is defined as the time from first response (CR or PR) to the date of initial objectively documented progression as determined using the 2014 Lugano classification or death due to any cause, whichever occurs first. For participants who neither progress nor die nor received

subsequent anti-cancer therapy, the DOR will be censored on the date of their last tumor assessment. Participants who start subsequent anti-cancer therapy without a prior reported progression will be censored at the last tumor assessment prior to or on the date of initiation of the subsequent anticancer therapy. The censoring scheme of DOR will be the same as the primary censoring scheme for PFS. This endpoint will only be evaluated in participants with objective response of CR or PR.

DOCR is defined as the time from first complete response (CR) to the date of initial objectively documented progression as determined using the 2014 Lugano classification or death due to any cause, whichever occurs first. The censoring scheme of DOCR will be the same as the primary censoring scheme for PFS. This endpoint will only be evaluated in participants with objective response of CR.

The other secondary efficacy endpoints include BOR, BOR, CRR, PR, TTR, DOR, and DOCR assessed by investigator.

4.2.2 Overall Survival

Overall Survival (OS): defined as the time between the date of randomization and the date of death. Participants who are alive at the time of analysis will be censored by last unknown alive date (LKAD).

4.2.3 Safety and Tolerability

Safety will be analyzed through the incidence of deaths, adverse events, serious adverse events, adverse events leading to discontinuation, adverse events leading to dose delay, select adverse events and specific laboratory abnormalities (worst grade) in each treatment group. Toxicities will be graded using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 4.0. See details in the Core Safety SAP.

Approved v3.0
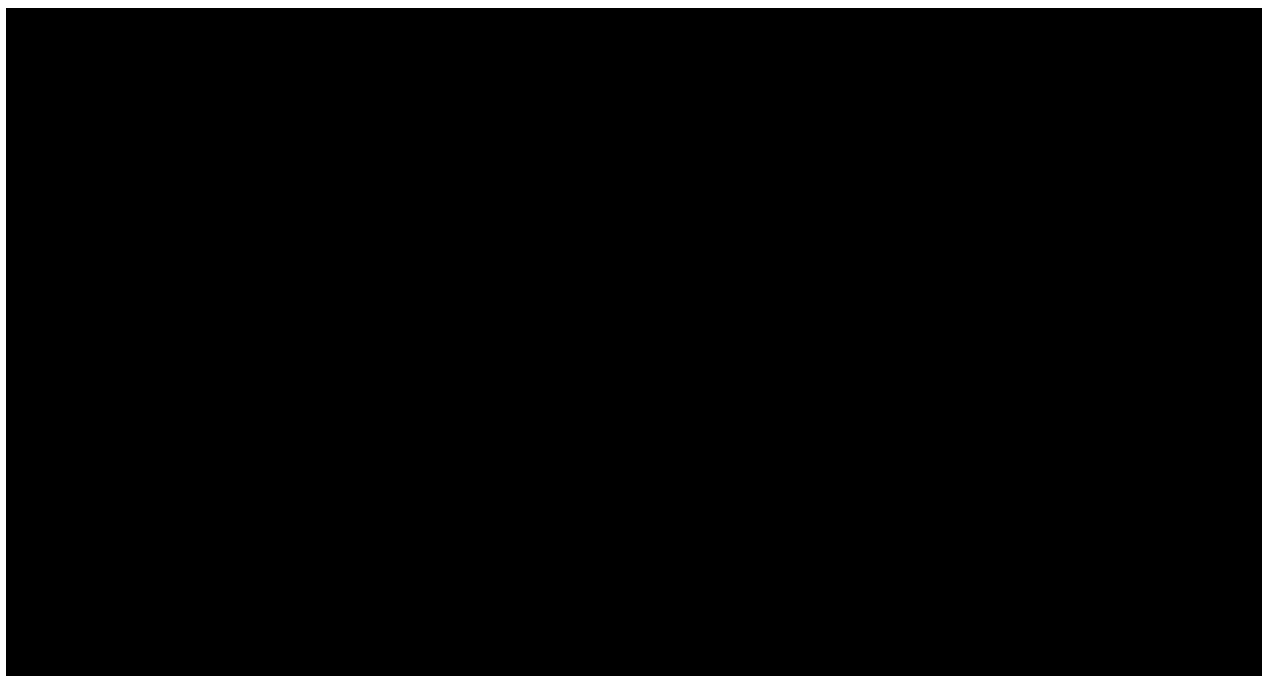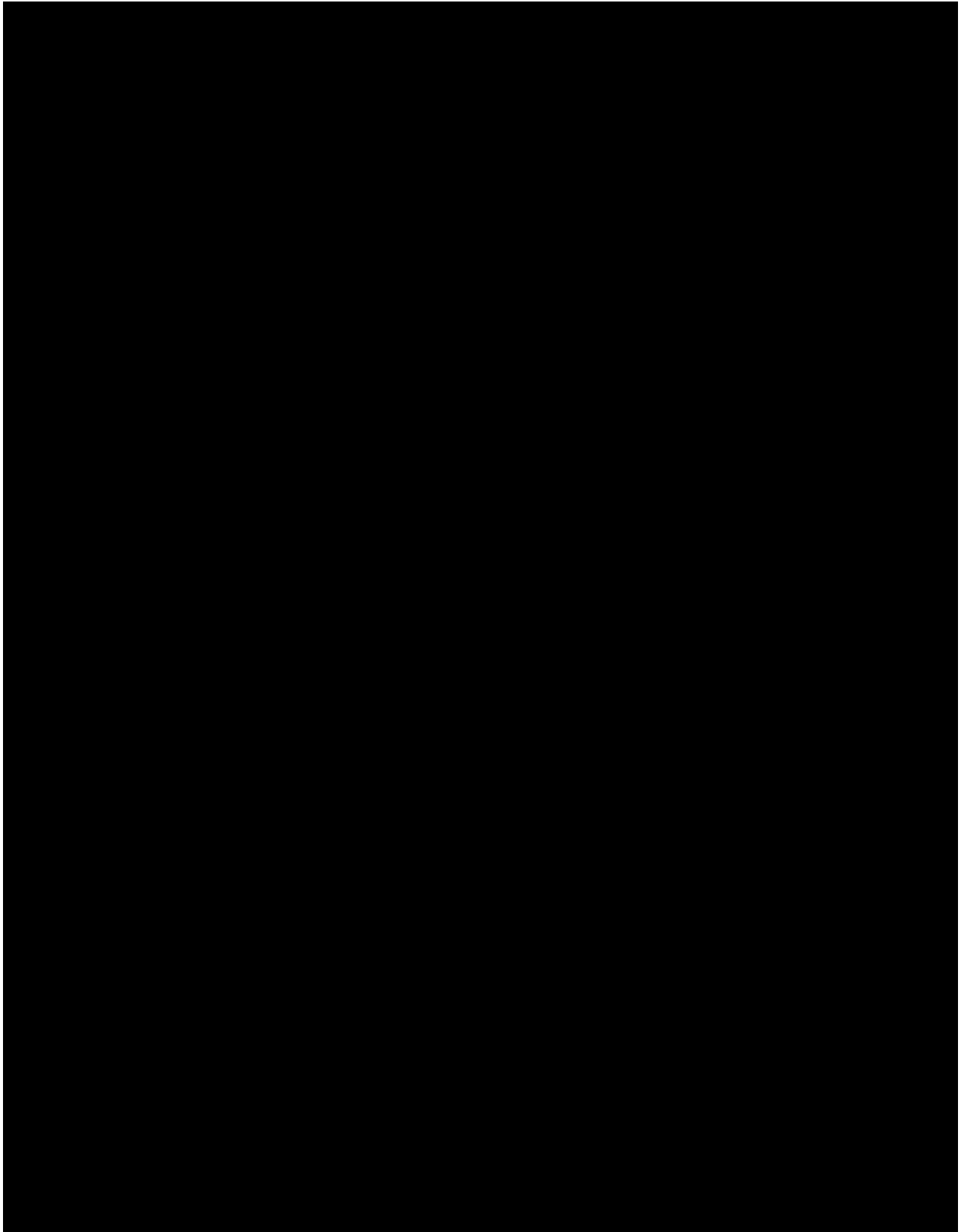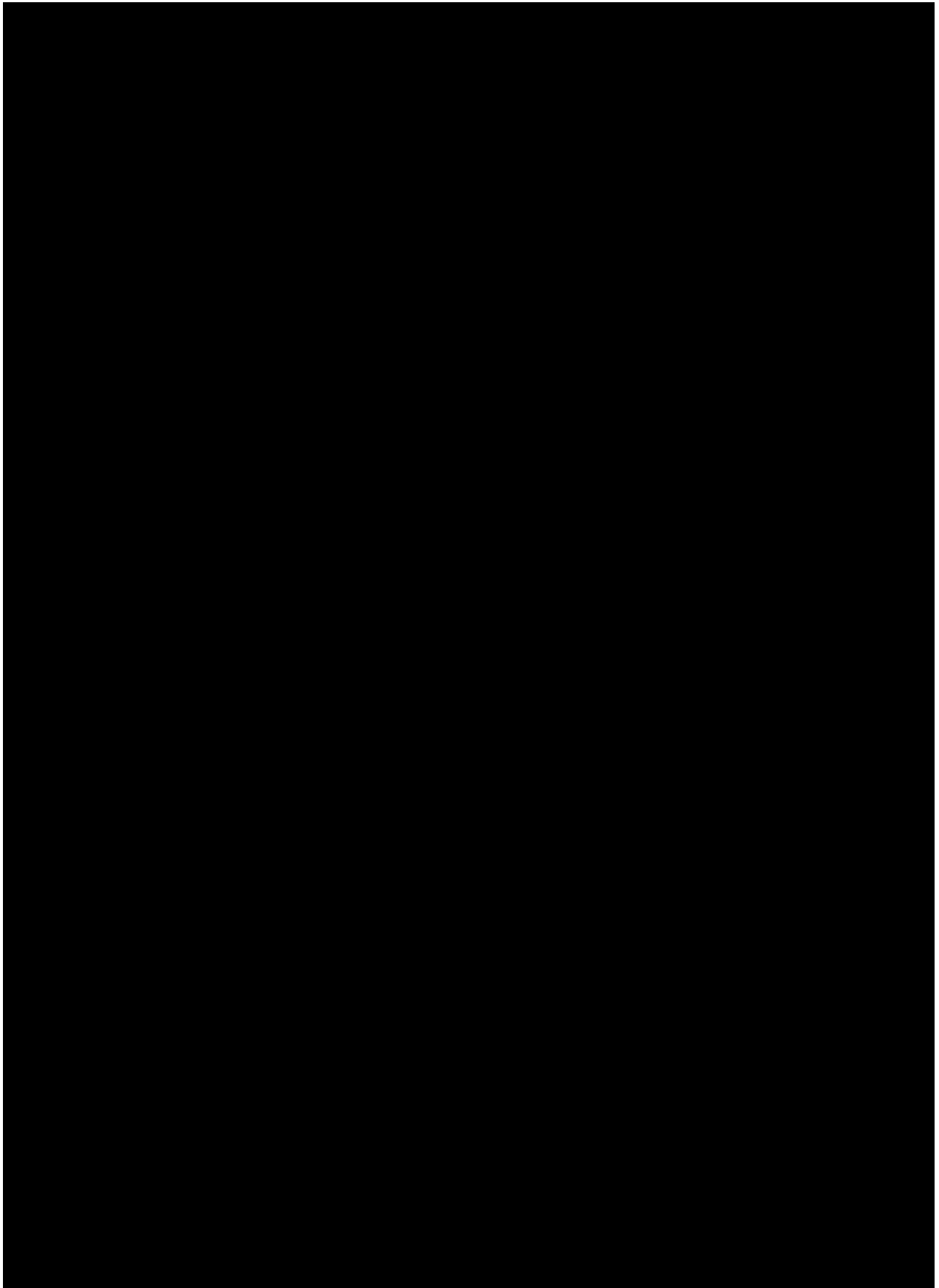
5   SAMPLE SIZE AND POWER

The planned sample size for this study will be approximately 340 randomized participants. The sample size of the study accounts for the primary efficacy endpoint, PFS. PFS will be evaluated for treatment effect at the overall alpha level of 0.05 (two-sided) with 90% power.
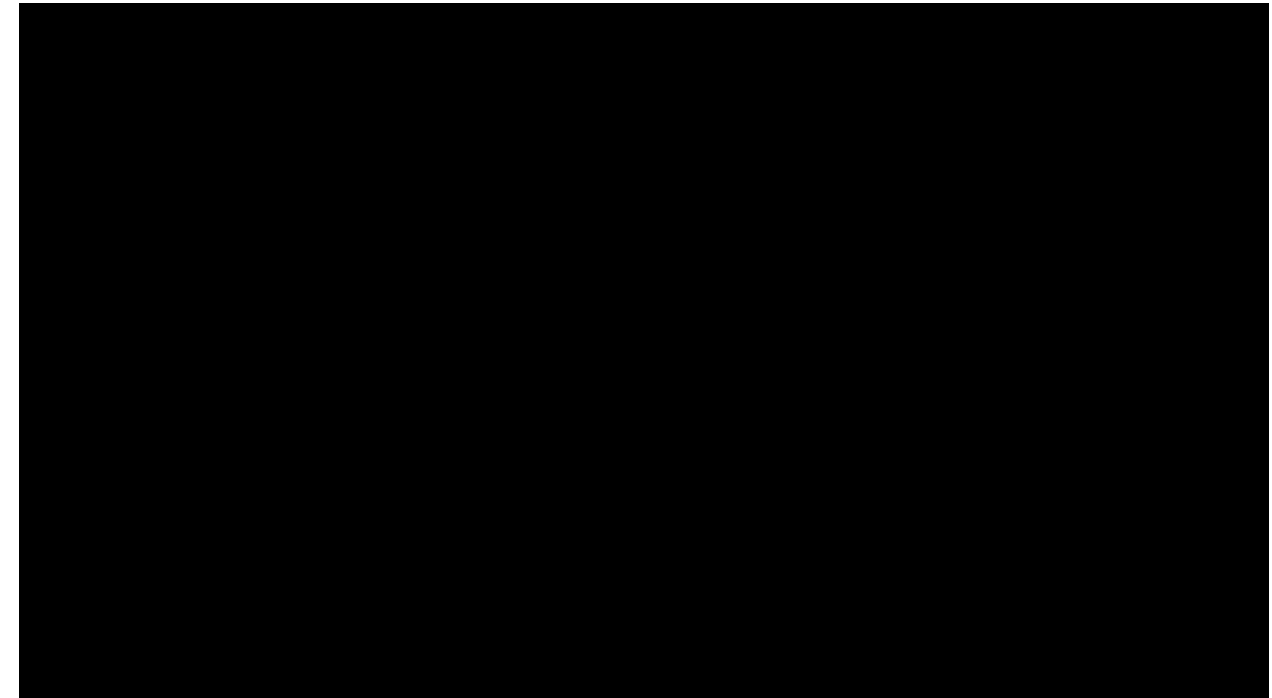
Table 5.1 summarizes the expected timing of each analysis. East version 6.3 was used for sample size/power computations.

Table 5.1 Schedule of Analyses

	Interim Analysis	Final Analysis
Conditions	At least 131 PFS events and 6 months follow-up	At least 187 events
Expected timing	40 months (34 months + 6 months of follow-up)	45 months (34 months+ 11 months of follow-up)
Alpha level	Interim PFS analysis projected at 0.0148 level [a]	Final PFS analysis projected at 0.0455 level

a. Using O'Brien and Fleming alpha spending function in case exact 131 PFS events are observed at the interim PFS analysis

6 STUDY PERIODS, TREATMENT REGIMENS AND POPULATIONS FOR ANALYSES

6.1 Study Periods

See Core Safety SAP.
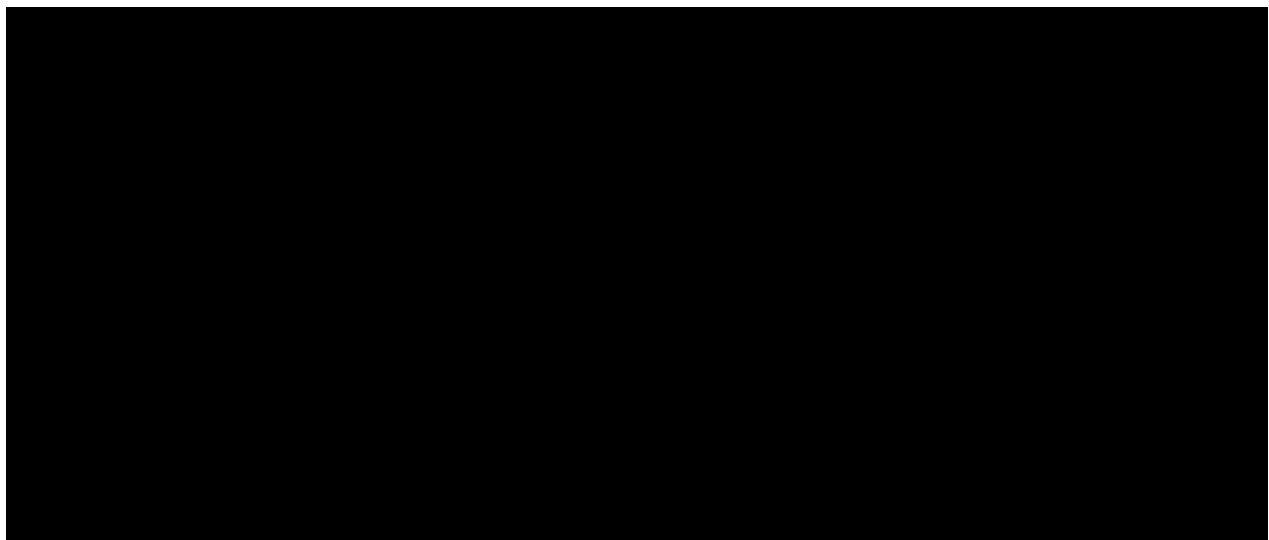
6.2 Treatment Regimens

The treatment group "**as randomized"** will be retrieved from the IVRS system
- Arm A: Experimental Arm nivolumab+BV.
- Arm B: Control arm BV.

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

6.3 Populations for Analyses

- **All enrolled participants**: All participants who signed an informed consent form and were registered into the IVRS. Analyses of the patients enrolled into the study but not randomized and the reason for not being randomized will be performed on the data set of all enrolled participants.
- **All randomized participants**: All participants who were randomized to any treatment group in the study. This is the primary dataset for analyses of demography, protocol deviations, baseline characteristics, efficacy, ███████████████████████.
- **All treated participants**: All participants who received at least one dose of study medication. This is the primary dataset for dosing and safety.
- **Response evaluable participants**: randomized participants whose change in the sum of diameters of target lesions was assessed (i.e.: target lesion measurements were made at baseline and at least one on-study tumor assessment).

7 STATISTICAL ANALYSES

7.1 General Methods

Unless otherwise noted, the bulleted titles in the following subsections describe tabulations of discrete variables, by the frequency and proportion of participants falling into each category, grouped by treatment (with total). Percentages given in these tables will be rounded and, therefore, may not always sum to 100%. Continuous variables will be summarized by treatment group (with total) using the mean, standard deviation, median, minimum and maximum values.

Time to event distribution (e.g. progression free survival, overall survival and duration of response) will be estimated using Kaplan Meier techniques.

Median survival time along with 95% CI will be constructed based on a log-log transformed CI for the survivor function S(t) [5,6]. Rates at fixed timepoints (e.g. PFS at 6 months) will be derived from the Kaplan Meier estimate and corresponding confidence interval will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function S(t)[7].

Unless otherwise specified, a stratified log-rank test will be performed to test the comparison between time to event distributions (e.g. PFS and OS). Stratification factors will be prior ASCT status (YES/NO) and prior BV use (YES/NO).

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.

The difference in rates between the two treatment groups along with their two-sided 95% CI will be estimated using the following Cochran-Mantel-Haenszel (CMH) method of weighting[8], adjusting for the stratification factors:

$$\hat{\theta} = \frac{\sum_i w_i \hat{\theta}_i}{\sum_i w_i} \sim N\left[\theta, \frac{\sum_i w_i^2 \left[\frac{p_{ix}(1-p_{ix})}{n_{ix}-1} + \frac{p_{iy}(1-p_{iy})}{n_{iy}-1}\right]}{\left(\sum_i w_i\right)^2}\right]$$

where $\hat{\theta} = p_{ix} - p_{iy}$ is the rate difference of the ith stratum, $w_i = \frac{n_{ix} n_{iy}}{n_{ix} + n_{iy}}$, and n_{ix} and n_{iy} are the number of participants randomized to treatments x and y, respectively, in the ith stratum.

Stratification factors will be same as above. Associated odds-ratio will be derived.

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

7.2 Study Conduct

7.2.1 Accrual

The accrual pattern will be summarized per country, investigational site and per month for all enrolled participants. Randomization date, first dosing date, country, investigational site will be presented in a by participant listing of accrual.

7.2.2 Relevant Protocol Deviations

The relevant Protocol Deviations will be only summarized for all randomized participants, by treatment group and overall. The following programmable deviations from inclusion and exclusion criteria will be considered as relevant protocol deviations. Non-programmable relevant eligibility and on-treatment protocol deviations, as well as significant (both programmable and non-programmable) eligibility and on-treatment protocol deviations will be reported through ClinSIGHT listings.

At entrance:
- Participants without either measurable disease at baseline or FDG avid by PET
- Participant with baseline ECOG >1

On-Study:
- Participants receiving concurrent anti-cancer therapy other than nivolumab or BV.
- Participant treated differently as randomized (participants who received the wrong treatment, excluding the never treated).

A summary table and a by participant listing of relevant protocol will be produced.

7.3 Study Population

7.3.1 Subject Disposition

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

Number of participants randomized but not treated along with the reason will be tabulated by treatment group as randomized. This analysis will be performed only on the all randomized population only.

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

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

7.3.2 Demographics and Other Baseline Characteristics

Descriptive statistics will be summarized the following baseline characteristics for all randomized participants by treatment group as randomized. All baseline presentations will identify participants with missing measurements.

- Age (descriptive statistics)
- Age categories (, <30, ≥30 and <45, ≥45 and <60, ≥60)
- Gender (male, female)
- Race (white, black, asian, other)
- Baseline ECOG Performance Status (0, 1)
- Weight (descriptive statistics)
- Time from initial disease diagnosis to first dose of study medication (< 3 months ; >=3 months)
- Prior ASCT status (YES/NO)
- Prior BV use (YES/NO).
- Region (US/Canada, Europe, Rest of the World)
- Smoking Status (Yes, No, Unknown)
- Disease stage at initial diagnosis (stage 1 or 2, 3, 4)
- International Prognostic Score (IPS) (0-2, 3 or higher) at initial diagnosis
- Disease stage at study entry (stage 1 or 2, 3, 4)
- B-symptoms at baseline (absent, present)
- Specific B-symptoms at baseline (Fever, Night Sweats, Weight Loss)
- Bulky disease at baseline (Yes, No, Not Available)
- Extra lymphatic involvement at baseline (Yes, No, Not Available)
- Lymphoma involvement in bone marrow at baseline (Yes, No, Not Available)
- All lesions (Investigator Tumor Assessments at Baseline): sites of diseases, number of disease sites per participant
- Target Lesions (Investigator Tumor Assessments at Baseline): Presence of target lesions, site of target lesion, sum of product diameters of target lesions.

7.3.3 Medical history- Concurrent diseases

General medical history will be listed by participant and pretreatment events will be tabulated.

7.3.4 Prior therapy agents

The following will be summarized:

- Number of participants by type of prior therapy received (excluding preparative regimen for ASCT): immunomodulatory derivatives (e.g. lenalidomide), other chemotherapy, immunotherapy by monoclonal antibodies (e.g. Rituximab, Brentuximab)

18

- Number of participants by type of regimen for first and second lines of therapy (e.g. ABVD, ICE, BEACOPP, Stanford V)
- Number of participants by type of regimen received for preparation to ASCT (BEAM, CBV, Other)
- Number of prior systemic regimens received (1, 2, 3, 4, ≥ 5), excluding preparative regimen before and after ASCT
- Best response to regimen before most recent ASCT
- Disease Status at most recent ASCT
- Best response to most recent ASCT
- Time from completion of most recent prior regimen to treatment (< 3, 3 - 6, > 6 months)
- Prior radiotherapy (yes or no)

Other Prior therapy:

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

Medication will be reported using the generic name. A listing by participant will also be provided.

7.3.5 Baseline examinations

Participants with abnormal baseline physical examination will be tabulated by examination criteria and by treatment group.

7.4   Extent of Exposure

Analyses in this section will be performed in all treated participants by three treatment groups as treated.

7.4.1 Administration of study therapy

The following parameters will be summarized (descriptive statistics) by drug (nivolumab or BV) and treatment group:

- Relative dose intensity (%) using the following categories: < 50%; 50 - < 70%; 70 - < 90%; 90 - < 110%; $\geq$ 110%.
- Number of doses received for each drug (summary statistics).
- Cumulative dose
- Number of cycles received.

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

A by-participant listing of dosing of study medication (record of study medication, infusion details, and dose change) and a listing of batch number will be also provided.

Below table summarizes the key parameters used to calculate dosing data.

Table 7.4.1-1:   **Administration of study therapy: definition of parameters**

	Nivolumab	BV
Dosing schedule per protocol	360mg every 3 weeks	1.8 mg/kg every 3 weeks
Dose	*Dose (mg)* is defined as Total Dose administered (mg). Dose administered in mg at each dosing date 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) is* sum of the doses (*mg*) administered to a participant during the treatment period.	*Cum dose (mg/kg) is* sum of the doses (*mg/kg*) administered to a participant during the treatment period.
Relative dose intensity (%)	Cum dose (mg) / [(Last dose date - Start dose date + 21) x 360 / 21] x 100	A: Cum dose (mg/kg) / [(Last dose date - Start dose date + 21) x 1.8 / 21] x 100
Duration of treatment	Last dose date - Start dose date +1	Last dose date - Start dose date +1

7.4.2 Modifications of Study Therapy

7.4.2.1 Dose delays

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 both nivolumab and BV. Participants may be dosed with nivolumab no less than 19 days from the previous dose.

The length of delay is defined as (duration of previous cycle in days - 21) for nivolumab or BV. Dose delays will be divided into following categories: on-time, 4-7 days, 8-14 days, 15-42 days, > 42 days. Reason for dose delay will be retrieved from CRF dosing pages.

The following parameters will be summarized by treatment group:
- Number of dose delayed per participant, Length of Delay and Reason for Dose Delay.

7.4.2.2 Infusion Modification

Each nivolumab or BV 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 drug:
- Number of participant with at least one dose infusion interrupted along with reason for the interruptions and number of infusions interrupted per participant.
- Number of participants with at least one infusion with IV rate reduced along with the reason of the rate reduction.

7.4.2.3 Dose Reductions

There will be no dose escalations or reductions of nivolumab allowed.

Only dose reduction of BV is permitted per protocol and should not be re-escalated without discussion with the sponsor.

Dose levels of BV are defined in the protocol as follows:
- Dose level 0.   1.8 mg/kg
- Dose level -1.   0.9 mg/kg

For any cycle (excluding Cycle 1), it will be defined as a dose reduction if the observed dose level (based on calculated administered dose) is below the dose level of the previously administered dose. Dose ranges for dose levels of BV are defined in Table 7.4.2.3-1.

Table 7.4.2.3-1:   **Calculated Dose Ranges and Related Dose Levels of Brentuximab**

Dose Range (mg/kg)	Dose Level
≥ 1.35	Level 0
< 1.35	Level -1

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.

The following will be summarized of dose reduction for all treatment groups:
- Number and percentage of participants with at least one dose reduction and reason of the dose reduction by drug and treatment groups,
- Number and percentage of participants with a dose reduction to dose level -1 by drug and treatment groups,

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 study therapy), will be coded using the WHO Drug Dictionary.

The following summary tables will be provided:
- Concomitant medications (participants with any concomitant medication, participants by medication class and generic term).

A by-participant listing will accompany the table.

7.4.4 Discontinuation of Study Therapy

The number and percentage of participants who have discontinued all study treatment and reason for discontinuation will be summarized by treatment group and overall using participant status eCRF page from end of treatment. This summary, unlike other safety analyses, will include all randomized participants and will be grouped by treatment group as randomized. This is done in order to give a full accounting of all participants who are off study treatment, including those who were randomized but never treated.

7.5 Efficacy

7.5.1 Primary analysis of PFS by BICR

The primary endpoint PFS assessed by BICR will be compared between the two randomized arms using a two-sided, log-rank test stratified by the same factors in the randomization. An O'Brien and Fleming α spending function will be employed to determine the significance levels and stopping boundaries at the interim look (overall alpha=0.05).

HR and corresponding two-sided 100 (1-α)% (α adjusted for the interim analysis) and 95% CI will be estimated in a stratified Cox proportional hazards model using randomized arm as a single covariate. The PFS curves for each treatment group will be estimated using the KM product-limit method. Two sided, 95% confidence intervals for median PFS will be constructed based on a log-log transformed CI for the survivor function S(t).

PFS rates at 6 months will be estimated using KM estimates on the PFS curve. PFS rates at 12, 18, 24, 36 months may also be estimated depending on whether minimum follow-up will be longer than or equal to timepoint to generate the rate. Associated two-sided 95% CIs will be calculated.

The source of PFS event (death vs. progression) will be summarized by treatment group. The status of participants who are censored in the PFS Kaplan-Meier analysis will be tabulated for each treatment group using following categories:

- Never treated
- On-study (on treatment or progression-free in follow-up).
- Off-study: (lost to follow-up, withdraw consent other).
- Received subsequent anti-cancer therapy.

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 (log 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. ████████████████████████████████
████████

7.5.1.1 Sensitivity analyses of PFS

Sensitivity analyses of PFS will also be performed as described below. Median PFS, HR (95% CIs), p -value from the log-rank test will be computed as in the primary PFS analysis

- *PFS accounting for assessment after subsequent therapy* participants will be defined similarly to the primary definition except that events (progression or death) and tumor assessments that occurred after subsequent anticancer therapy will be taken into account. A Kaplan-Meier plot will be produced

Table 7.5.1.1-1: **Censoring scheme 1 for Sensitivity Analysis of PFS**

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization date	Censored
No on-study tumor assessments and no death	Randomization date	Censored
New subsequent anticancer treatment started without a prior reported progression by BICR or death	Date of last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored
Progression per BICR documented at a scheduled or unscheduled visit and no other subsequent anticancer treatment started before progression	Date of the first documented tumor progression	Progressed
Participant progression free, no death and no subsequent anti-cancer therapy treatment started	Date of last evaluable tumor assessment	Censored
Death without prior progression per BICR	*Date of death*	*Progressed*

- *PFS accounting for missing tumor assessment prior to PFS event (progression or death)*. This analysis will be performed only if at least 20% of events have missing prior tumor assessment. It will apply the following restriction to the primary definition: If the elapsed time between the PFS event and the last on-study tumor assessment immediately prior to the event (or randomization date if no on-study scan) is two or more missed visits (more than 12 weeks + 10 days), the participant will be censored at his last tumor assessment prior to the PFS event (or randomization date if no on-study scan). A Kaplan-Meier plot will be produced.

Table 7.5.1.1-2: **Censoring scheme 2 for Sensitivity Analysis of PFS**

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization date	Censored
No on-study tumor assessments and no death	Randomization date	Censored
New subsequent anticancer treatment started without a prior reported progression by BICR or death	Date of last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy	Censored
Progression per BICR documented at a scheduled or unscheduled visit and no other subsequent anticancer treatment started before progression	Date of the first documented tumor progression	Progressed
Participant progression free, no death and no subsequent anti-cancer treatment started	Date of last evaluable tumor assessment	Censored
Death without prior progression per BICR and no subsequent anti-cancer therapy treatment started	Date of death	Progressed

Death or progression after two or more missed visits	*Date of last tumor assessment prior to the PFS event*	*Censored*

7.5.1.2 Consistency of treatment effect on PFS in subsets

To assess consistency of treatment effect on PFS in different subsets, a "forest" plot of the PFS un-stratified hazard ratio (and 95% CI) will be produced for the following variables:

- Prior ASCT status (YES/NO)
- Prior BV use (YES/NO).
- Disease Stage at initial diagnosis (stage 1 or 2, 3, 4)
- Region (US/Canada, Europe, Rest of World).
- Age categorization (<30, ≥30 and <45, ≥45 and <65, , ≥65 and <75, ≥75 and <85, ≥ 85)
- Gender (Male, Female).
- Race (White, African American, Asian, and Other).
- ECOG PS (0 vs. 1).
- Smoking status (yes, no, unknown)
- Time from the initial diagnosis to the randomization date (<3 months (yes vs.no))

If a subset category has less than 10 participants (after collapse) per treatment group, HR will not be computed/displayed. Number of events and median PFS along with 95% CI will be displayed for each randomized treatment group.

7.5.1.3 Multivariate PFS analysis

A multivariate analysis using a stratified Cox model will be fitted to assess the treatment effect on OS when adjusted for potential prognostic factors. The stratification factors in the model are the same as the ones in the randomization. The following prognostic factors (source: CRF) will be included in the model in addition to treatment group variable:

- Time from diagnosis to start of the randomization (< 3 months (yes vs. other)).
- Age categorization (< 30, ≥ 30).
- Gender (Male vs. Female).
- Baseline ECOG PS (0 vs. 1).

HR and 95% CI will be provided for treatment variable and all covariates. Descriptive p-values will be provided.

7.5.2 Secondary Efficacy Endpoints

7.5.2.1 CRR

The secondary endpoint CRR based on BICR assessment will be compared in two randomized arms via a two-sided, Cochran-Mantel-Haenszel (CMH) test at 0.05 level stratified by the same factors used in randomization. An estimate of the treatment odds ratio between the nivolumab+BV

and BV arm along with corresponding two-sided 95% CIs and p-value will be presented. An estimate of the difference in CRRs and corresponding 95% CI will be calculated using CMH methodology and adjusted by the same stratification factors.

The estimate of the response rates and associated exact two-sided 95% CIs (by Clopper and Pearson method) will be presented for both arms.

The CRR analysis will be performed at the time of interim and final PFS analysis. A hierarchical testing procedure is employed to control the study-wise type I error at 5%. For more details on the testing procedure, please refer to the interim analysis section 7.5.4.

To assess consistency of treatment effect on CRR in different subsets, a "forest" plot of the unweighted differences in CRR and corresponding exact 95% CI using the method of Newcombe approach will be produced for the same subsets as defined in the primary PFS analysis.

If a subset category has less than 10 participants (after collapse) per treatment group, the differences in CRR will not be computed/displayed.

7.5.2.2 BOR, ORR, TTR, DOR and DOCR

The Best Overall Response (BOR) based on BICR assessment (using 2014 Lugano Classification) will be summarized by response category for each treatment group. ORR as defined as proportion of participants who achieved CR or PR before taking any subsequent anti-cancer therapy, will be computed in each treatment group along with the exact 95% CI using Clopper-Pearson method.

An estimate of the difference in ORRs and Odds ratio and corresponding 95% CIs will be calculated using CMH methodology and adjusted by the same stratification factors as in primary analysis of PFS. A by participant listing of BOR and Tumor Measurements will be provided.

To assess consistency of treatment effect on ORR in different subsets, ORR will be computed across the same subsets as defined in the PFS analysis.

Summary statistics of time to objective response will be provided for each treatment group for participants who achieve PR or CR, as assessed by BICR. To assess tumor response kinetics, time to response will also be analyzed using the KM methodology for all treated participants. Kaplan-Meier curve will represent the cumulative rate of response over time. For the non-responders, time to response will be censored at the maximum time to response + 1 day of all participants in their respective treatment group. Cumulative response rates will be tabulated for Month 6, 12, 18 and 24.

Duration of response in each treatment group will be estimated using KM product-limit method for participants who achieve PR or CR. Median values along with two-sided 95% CI will be calculated. Duration of complete response will be analyzed in a similar manner.

The following participant-level graphics will also be provided by treatment group as randomized:
- For the responders only, time courses of the following events of interest will be graphically displayed: tumor response, tumor progression, last dose received, and death.
- For response evaluable participants, a waterfall plot showing the best reduction in target lesion will be produced.

The CRR, ORR, BOR, DOR and DOCR assessed by investigator will be analyzed similarly to the same endpoints assessed by BICR.

7.5.2.3 OS

The distribution of OS will be compared in two randomized arms via a two-sided, log-rank test stratified (per IVRS) by the same factors in the randomization. The hazard ratio (HR) and the corresponding 95% CI will be estimated in a stratified Cox proportional hazards model using randomized arm as a single covariate. The stratification factors in the model will be the same as the ones in the primary PFS analysis.

The OS curves for each treatment group will be estimated using the Kaplan-Meier (KM) product-limit method. Two-sided, 95% confidence intervals for median OS will be constructed based on a log-log transformed CI for the survivor function S(t).

Survival rates at 6, 12, 18, 24, 36 months may also be estimated using KM estimates on the OS curve for each randomized arm. Minimum follow-up must be ≥ timepoint to generate the rate. Associated two-sided 95% CIs will be calculated using the Greenwood's formula for variance derivation and on log-log transformation applied on the survivor function S(t).

The status of participants who are censored in the OS Kaplan-Meier analysis will be tabulated for each treatment group using following categories:
- On-study (on-treatment or in follow-up).
- Off-study (lost to follow-up, withdraw consent, etc.).

To assess consistency of treatment effects in different subsets, a "forest" plot of the OS un-stratified hazard ratio (and 95% CI) will be produced for the same subsets as defined in the primary PFS analysis.

7.5.3 Efficacy Analyses for Both PFS and OS

7.5.3.1 Participant Follow-up

The minimum follow-up will be reported. The minimum follow-up is defined as the time interval between the last patient's randomization date and the clinical cutoff date.

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

The currentness of follow-up, defined as the time between last OS contact (i.e., last known date alive or death date) and data cut-off date, will be summarized by treatment group. Participants who died before data cut-off date will automatically have zero value for currentness of follow-up. For participants with last known date alive after data cut-off date, they will have zero value for currentness of follow-up as well. The currentness of follow-up will be categorized into the following categories: 0 days, 1-3 months, 3-6 months, 6-9 months, 9-12 months and ≥ 12 months.

7.5.3.2 Follow-Up Therapy

The following information pertaining to subsequent therapies will be summarized:

Number and percentage of participants receiving subsequent therapies including:
- Chemotherapy by drug name.
- Immunotherapy (anti-PD1 agents, anti-CTLA4 agents and others, by drug name).
- Brentuximab
- Other investigational agent by drug name.
- Transplant/Surgery.
- Radiotherapy.
- Any combination of the above.

A participant listing of follow-up therapy will be produced for participants who had any subsequent therapy.

7.5.4 Interim Analysis

An independent statistician external to BMS will perform the analysis. In addition to the formal planned analysis on CRR and interim analysis for PFS, the DMC will have access to periodic unblinded interim reports of efficacy and safety to allow a risk/benefit assessment. Details are included in the DMC charter.

A formal interim analysis for superiority of PFS in participants randomized to receive nivolumab+BV vs. participants randomized to receive BV will be performed on all randomized participants when at least 131 PFS events have been observed (approximately 70% (131/187) of the total number of events required for the final PFS analysis).

This PFS comparison will be tested using the interim monitoring feature of East software (version 6.3) based on a generalization of the Lan-DeMets error spending function approach using an O'Brien-Fleming stopping boundary[9] to reject H_0, ($H_0 =$ no treatment difference) controlling for a two-sided overall alpha of 5.0%. For example, if exactly 131 PFS events were in the locked database for the interim analysis, H_0 would be rejected at the interim if the p-value from the stratified log-rank test is $p < 0.0148$. If the study continues beyond the interim look, H_0 would be rejected at the final PFS analysis if the p-value from the stratified log-rank test is $p < 0.0455$.

If the number of PFS events is not exactly 131 at the time of the interim analysis, the nominal critical points and values for interim and final analyses will be calculated based upon the observed information fraction (number of events at the interim over the planned 187 events at the final).

A hierarchical testing procedure is employed to control the study-wise type I error at 5% level. Specifically, only if the primary PFS analysis result is significant at either interim or final analyses, the comparison of CRR between two treatment arms will be performed at an alpha level of 5%. Otherwise, CRR will not be compared between two arms, and consequently the p-value will not be reported for CRR. In that case, only CRR in both treatment arms will be reported with 95% CI.

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

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

Implications of PFS Interim Analysis

At the time of the formal interim analysis for superiority of PFS, the DMC may recommend continuing or stopping the trial.

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

If the trial is stopped for superiority of PFS at the interim, the p-value from the interim stratified log-rank test will be considered the final primary PFS analysis result. The p-values from these analyses will be considered as the final results.

7.6 Safety

7.6.1 Deaths

See Core Safety SAP.

7.6.2 Serious Adverse Events

See Core Safety SAP.

7.6.3 Adverse Events Leading to Discontinuation of Study Therapy

See Core Safety SAP.

7.6.4 Adverse Events Leading to Dose Delay of Study Therapy

See Core Safety SAP.

7.6.5 Adverse Events

See Core Safety SAP.

7.6.6 Multiple Events

See Core Safety SAP.

7.6.7 Select Adverse Events

See Core Safety SAP.

7.6.8 Clinical laboratory evaluations

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

7.6.8.1 Hematology

See Core Safety SAP.

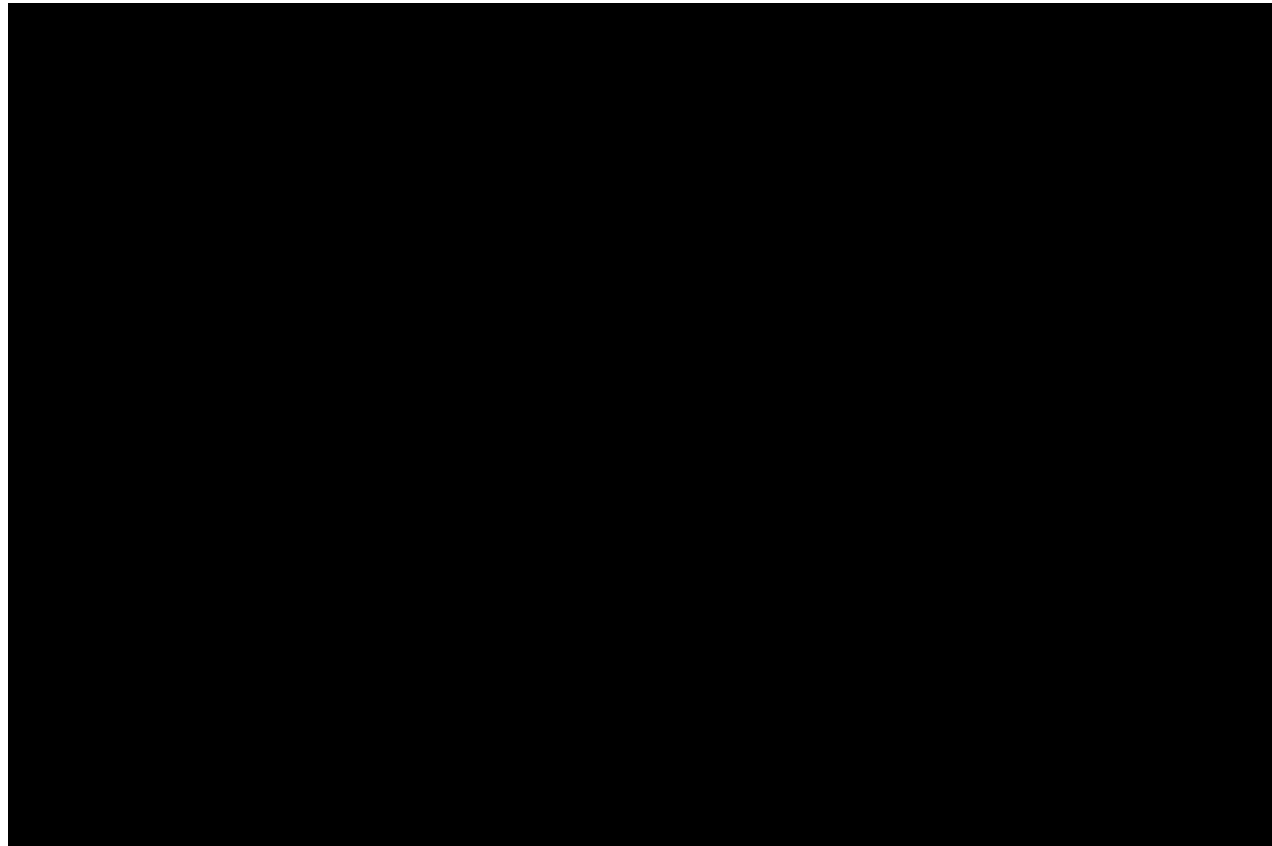### 7.6.8.2 Serum Chemistry

See Core Safety SAP.
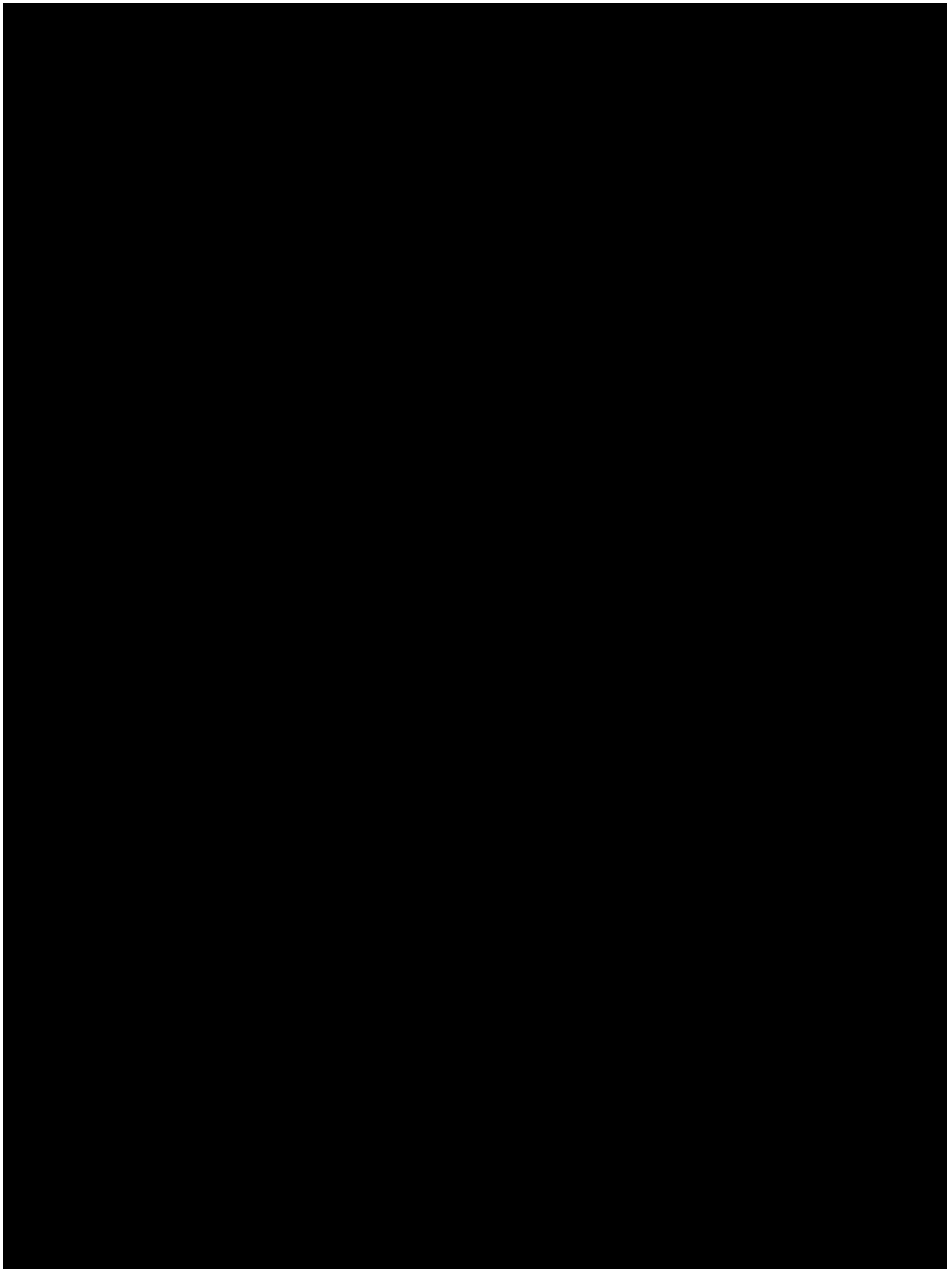
7.6.9 Vital Signs and Pulse Oximetry
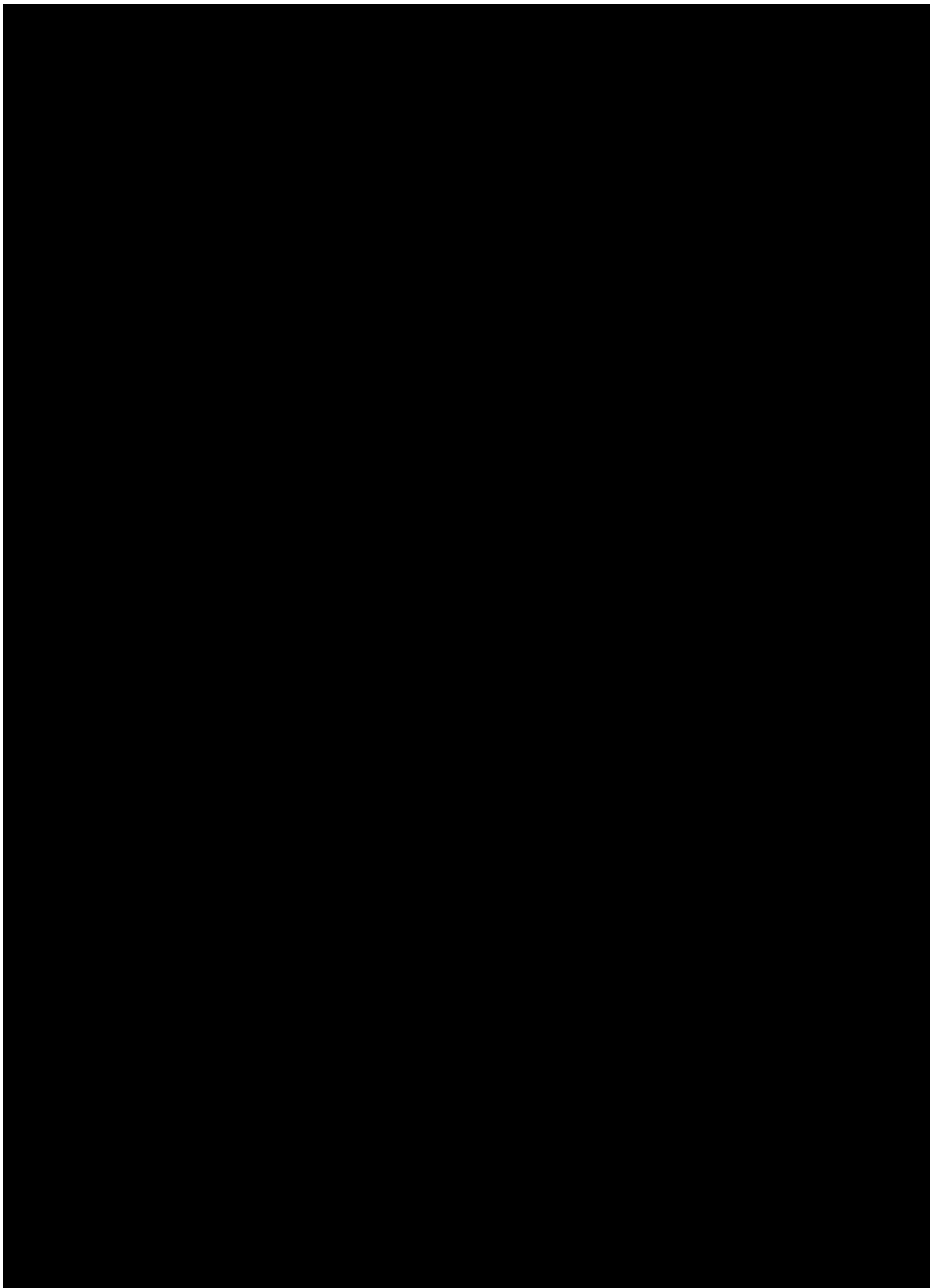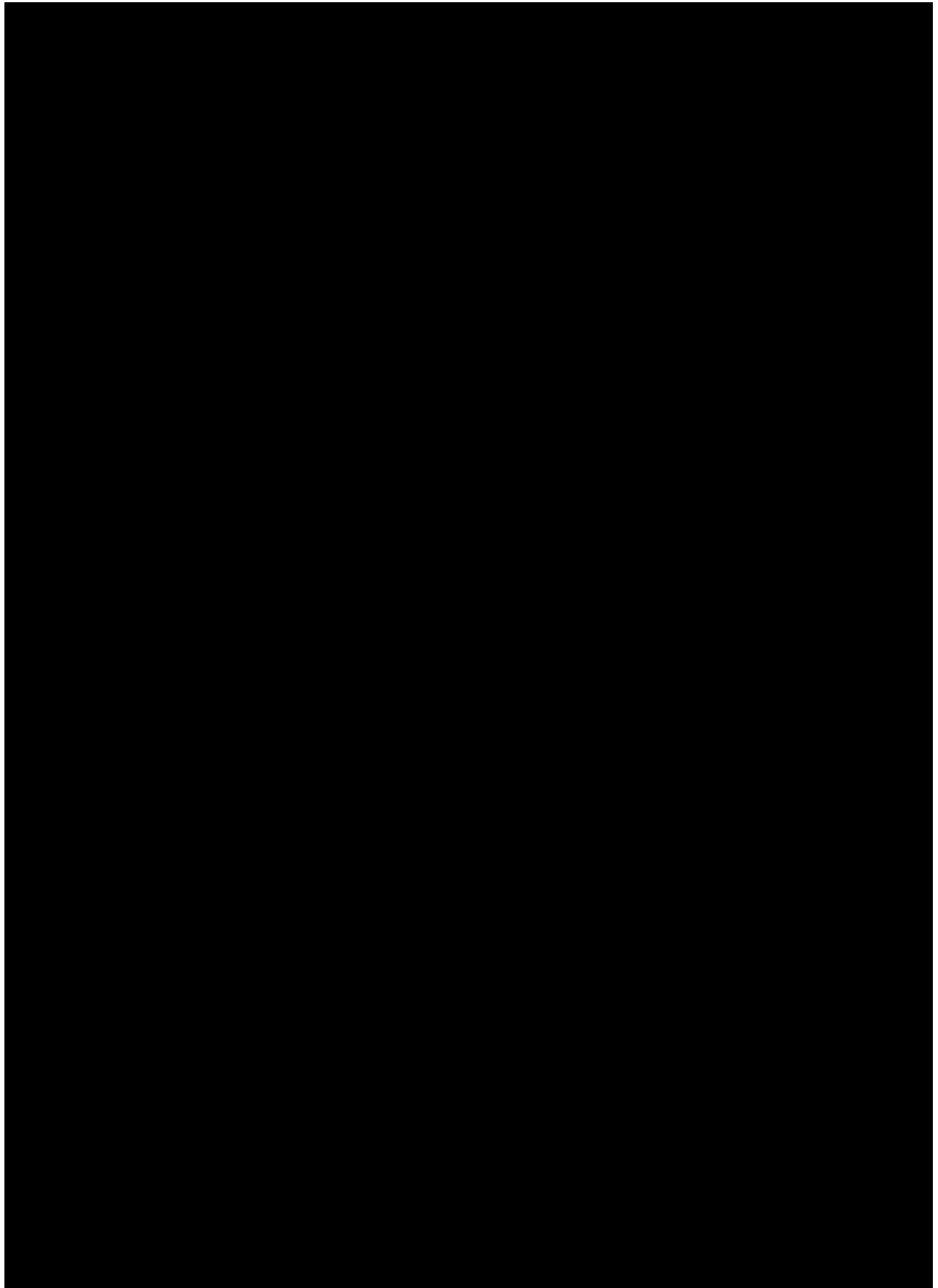
See Core Safety SAP.

7.6.11 Pregnancy

See Core Safety SAP.

Approved v3.0

Approved v 3.0

8 CONVENTIONS

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

For missing and partial adverse event onset dates, imputation will be performed using the Adverse Event Domain Requirements Specification[10]. Missing and partial Non-Study Medication Domain dates will be imputed using the derivation algorithm described in BMS Non-Study Medication Domain Requirements Specification[11].

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 alive date.

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

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

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

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

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

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

$$1 \text{ month} = 30.4375 \text{ days and } 1 \text{ year} = 365.25 \text{ days}$$

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

$$\text{Duration} = (\text{Last date - first date} + 1)$$

All statistical analyses will be carried out using SAS (Statistical Analysis System software, SAS Institute, North Carolina, USA) unless otherwise noted.
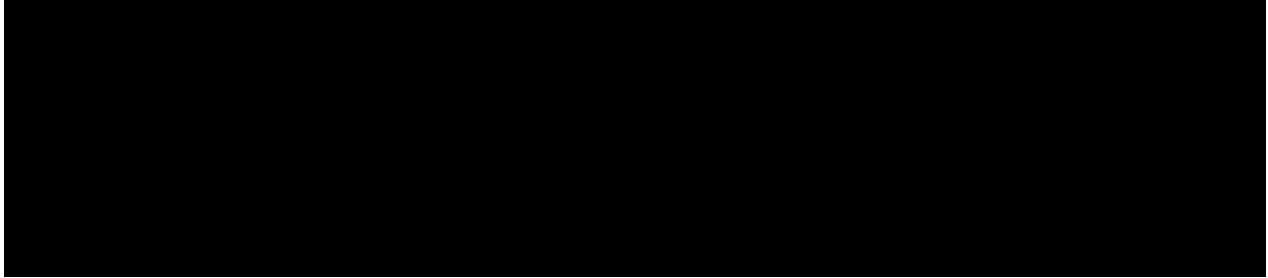
Safety conventions from Programming may be summarized separately in an appendix.

9   CONTENT OF REPORTS

All analyses describe in this SAP will be included in the final Clinical Study Report. Refer to the Data Presentation Plan for mock-ups of all tables and listings.

10 REFERENCES

1. Core Safety Statistical Analysis Plan for CA209, Bristol-Myers Squibb.

5. Brookmeyer R. and Crowley J (1982). A Confidence Interval for the Median Survival Time. Biometrics 38:29-41,

6. Klein, J. P. and Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censored and Truncated Data, page 120, New York: Springer-Verlag, 2nd Edition.

7. Kalbfleisch RL (2002). The Statistical Analysis of Failure Time Data. Section 1.4.1, 14-19, Wiley-Interscience, 2nd Edition.

8. Statistical Methodology in the Pharmaceutical Sciences (1990). Berry DA Ed., Chapter 13 Categorical Data analysis, 415 -417, Marcel Dekker.

9. O'Brien P.C. and Fleming T.R. (1979). A multiple testing procedure for clinical trials. Biometrics, vol. 35 no 3: 549-556.

10. Adverse Event Domain Requirements Specification. Bristol Myers Squibb Co. PRI. Version 2.1. April 23, 2012.

11. Non-Study Medication Domain Requirements Specification. Bristol Myers Squibb Co. PRI. Version 2.2 April 24, 2012.

11 DOCUMENT HISTORY

Table 11-1:   **Document History**

Version Number	Author(s)	Description
1.0	██████	Initial version- 24-Feb-17