Statistical Analysis Plan

A Phase 1/2, Open-label Study to Evaluate the Safety and Antitumor Activity of MEDI0680 (AMP-514) in Combination with Durvalumab versus Nivolumab Monotherapy in Subjects with Select Advanced Malignancies

Protocol Number: D6020C00001

Table of Contents

List of In-Text Tables

List of In-Text Figures

List of Abbreviations

Abbreviation	Definition
AE	adverse event
AESI	adverse event of special interest
AZDD	Astrazeneca Drug Dictionary
BOR	best overall response
CI	confidence interval
CR	complete response
DC	disease control
DCR24	disease control rate (SD maintained ≥ 24 weeks)
DCR8	disease control rate (SD maintained ≥ 8 weeks)
DLT	dose-limiting toxicity
DR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organization for Research and Treatment of Cancer
irCR	immune-related complete response
irNE	immune-related not evaluable
irPD	immune-related progressive disease
irPR	immune-related partial response
irRECIST	immune-related Response Evaluation Criteria in Solid Tumors
irSD	immune response stable disease
IV	intravenous
IXRS	interactive voice/web response system
MedDRA	Medical Dictionary for Regulatory Activities
MTD	maximum tolerated dose
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NE	not evaluable
NSCLC	non-small cell lung cancer
OR	objective response
ORR	objective response rate
OS	overall survival
PD	progressive disease
PD-1	programmed cell death 1
PD-L1	programmed cell death ligand 1
PFS	progression-free survival
PK	pharmacokinetics
PK-PD	pharmacokinetics - pharmacodynamics
PR	partial response
PRO	patient-reported outcome
Q2W	every two weeks

Abbreviation	Definition
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SD	stable disease
SPP	statistical programming plan
TDA	time to disease assessments
TEAE	treatment-emergent adverse event
TTR	time to response

1 INTRODUCTION

This document describes the statistical analysis for Study D6020C00001, a multicenter, open-label, Phase 1/2 study to evaluate the safety, tolerability, PK, immunogenicity, and antitumor activity of MEDI0680 in combination with durvalumab or nivolumab monotherapy in adult immunotherapy-naïve subjects with selected advanced malignancies. As background information, an overview of the study design is provided. The main portion of the document details the statistical analyses relating to each study objective as well as the general conventions and definitions that will be used. In addition, a set of table templates and specifications will be included in a statistical programming plan (SPP) to complement this document.

2 STUDY OVERVIEW

2.1 Study Objectives

2.1.1 Primary Study Objective

Dose-escalation Phase

To determine the maximum tolerated dose (MTD) or the highest protocol-defined dose in the absence of exceeding the MTD for MEDI0680 in combination with durvalumab and the safety profile of MEDI0680 in combination with durvalumab in subjects with advanced malignancies

Dose-expansion Phase

To evaluate the antitumor activity of MEDI0680 in combination with durvalumab versus nivolumab monotherapy in immunotherapy-naïve subjects with advanced or metastatic ccRCC as based on investigator assessed response using Response Evaluation Criteria in

Solid Tumors Version 1.1 (RECIST 1.1) (Eisenhauer et al, 2009).

2.1.2 Secondary Study Objectives

Dose-escalation:

To evaluate the antitumor activity of MEDI0680 in combination with durvalumab in subjects with advanced malignancies as based on the investigator assessed response using modified RECIST 1.1

Dose-expansion:

1. To describe the safety and tolerability of MEDI0680 in combination with durvalumab or nivolumab monotherapy in immunotherapy-naïve subjects with advanced or metastatic ccRCC
2. To evaluate the antitumor activity of MEDI0680 in combination with durvalumab versus nivolumab monotherapy in immunotherapy-naïve subjects with advanced or metastatic ccRCC as based on blinded independent central review (BICR) assessed response using RECIST 1.1

Dose-escalation and Dose-expansion:

1. To describe the pharmacokinetics (PK) of MEDI0680 in combination with durvalumab
2. To describe the PK of durvalumab in combination with MEDI0680
3. To determine the immunogenicity of MEDI0680 in combination with durvalumab
4. To determine the immunogenicity of durvalumab in combination with MEDI0680
5. To determine whether PD-L1 is a predictive biomarker for response to therapy with MEDI0680 in combination with durvalumab

2.1.3 Exploratory Study Objectives

1. To identify biomarkers that are predictive of antitumor response to MEDI0680 in combination with durvalumab
2. To profile gene expression changes that may correlate with antitumor response to MEDI0680 in combination with durvalumab
3. To evaluate additional biomarkers that may correlate with antitumor activity of MEDI0680 in combination with durvalumab compared with nivolumab monotherapy

4. To evaluate the pharmacodynamic activity of MEDI0680 in combination with durvalumab versus nivolumab monotherapy in the periphery.
5. To compare the pharmacodynamic changes resulting from complete PD-1/PD-L1 pathway blockade versus treatment with nivolumab monotherapy
6. To evaluate the antitumor activity of MEDI0680 in combination with durvalumab versus nivolumab monotherapy in subjects with advanced or metastatic ccRCC as assessed by immune-related Response Evaluation Criteria in Solid Tumors (irRECIST)

2.2 Study Design

This is a multicenter, open-label, Phase 1/2 study to evaluate the safety, tolerability, PK, immunogenicity, and antitumor activity of MEDI0680 in combination with durvalumab or nivolumab monotherapy in adult immunotherapy-naïve subjects with selected advanced malignancies.

The study includes 2 phases, dose-escalation and dose-expansion. In the dose-escalation phase, subjects with selected solid tumors will receive MEDI0680/durvalumab combination therapy. In the dose-expansion phase, subjects with ccRCC will receive either MEDI0680/durvalumab combination therapy or nivolumab monotherapy.

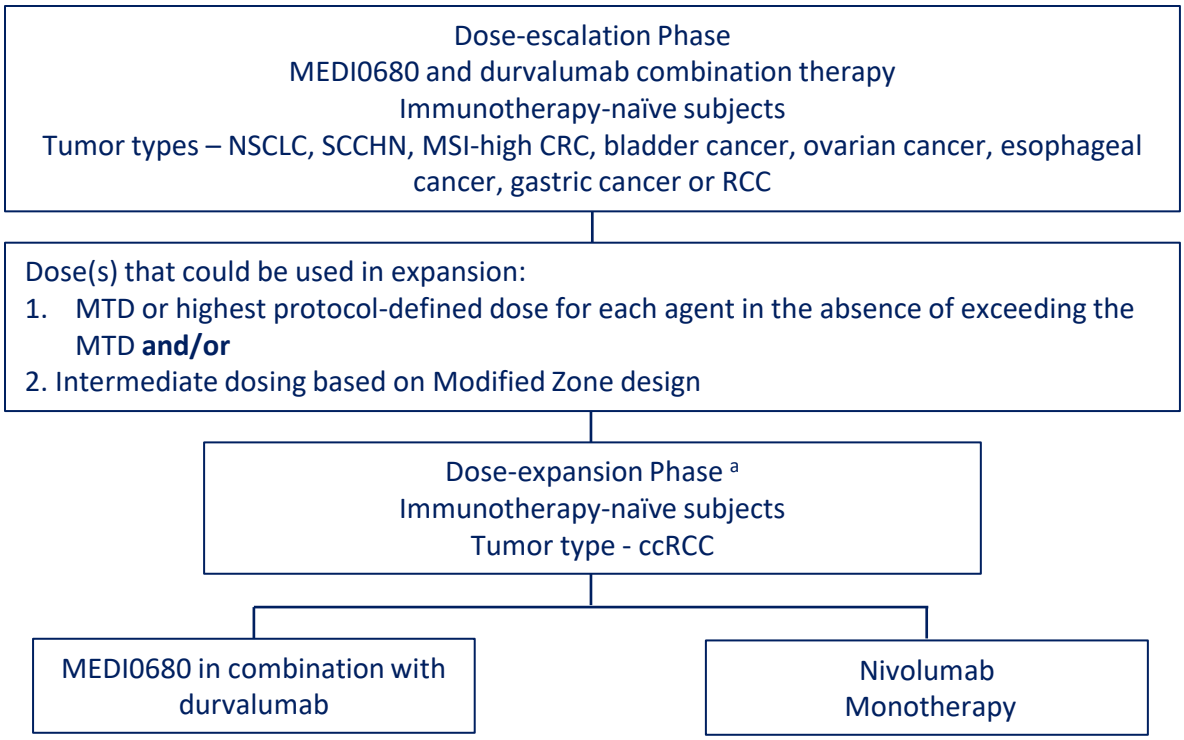
In the dose-escalation phase, subjects may receive treatment for up to 12 months; in the dose expansion phase, subjects may remain on treatment until unacceptable toxicity, confirmed progressive disease (PD), or development of other reason for treatment discontinuation. Subjects in dose expansion may receive investigational product(s) (IP) for a maximum of 2 years. All subjects will be evaluated regularly. Clinical status will be classified according to modified RECIST 1.1 for subjects in the dose-escalation phase and by RECIST 1.1 for subjects in the dose expansion phase, and other disease-specific assessments. All subjects will be followed for survival until the end of study. Adverse events (AEs) and serious AEs (SAEs) will be followed.

In the dose-escalation phase, subjects will be enrolled using a 3+3 dose escalation design in 1 of 6 dose-level treatment groups: 0.1 mg/kg MEDI0680 Q2W with 3 mg/kg durvalumab Q2W, and increasing dose levels of MEDI0680 Q2W (0.1, 0.5, 2.5, 10, and 20 mg/kg) with 10 mg/kg durvalumab Q2W. The dose-escalation phase will initially include immunotherapy-naïve subjects with solid tumors. As of Protocol Amendment 2, only subjects with non-small cell lung cancer (NSCLC), squamous cell carcinoma of the head and neck (SCCHN), microsatellite instability-high (MSI high) colorectal cancer (CRC), bladder

cancer, ovarian cancer, esophageal cancer, gastric cancer, or renal cell carcinoma (RCC) will be enrolled.

The dose-expansion phase will begin after the dose-escalation phase. Immunotherapy-naïve subjects with ccRCC will be randomized in a 2:1 ratio to either (1) MEDI0680/durvalumab combination therapy or (2) nivolumab monotherapy. Stratification factors will include the Memorial Sloan Kettering Cancer Center (MSKCC) risk group (0 = favorable risk; 1 or 2 = intermediate risk; 3 = poor risk) and PD-L1 expression status (≤ 1% and > 1%). Up to 40 subjects may be randomized in to the MEDI0680/durvalumab combination therapy arm and up to 20 subjects in the nivolumab monotherapy arm based on evaluation of emerging safety and efficacy parameters in the current study as well as other ongoing studies. Interim futility analyses will be performed in a continuous manner using Bayesian Predictive Probability (see Section 4 Interim Analysis).

Figure 2-1 Study Design



ccRCC = clear-cell RCC; CRC = colorectal cancer; MSI = microsatellite instability; MTD = maximum tolerated dose; NSCLC = non-small cell lung cancer; RCC = renal cell carcinoma; SCCHN = squamous cell carcinoma of head and neck.

[a] Up to a total of 60 subjects with ccRCC will be randomized in a 2:1 ratio to receive either MEDI0680/durvalumab combination therapy or nivolumab monotherapy.

2.2.1 Treatment Regimen

Subjects will be treated in either the dose-escalation or dose-expansion phase. MEDI0680, durvalumab and nivolumab will be administered via IV infusion. In dose-escalation phase, subjects will remain on treatment for up to 12 months. In the dose-expansion phase, subjects will remain on treatment until unacceptable toxicity, confirmed PD, or development of other reason for treatment discontinuation.

In the dose-escalation phase only, subjects who achieve and maintain DC (ie, CR, PR, or stable disease [SD]) through the end of the 12-month treatment period will enter follow-up. During the first 12 months of follow-up, if the subject has PD, the subject may be re-administered MEDI0680 and durvalumab for up to another 12 months with the same treatment guidelines followed during the initial 12-month period if the subject fulfills the criteria for re-treatment in the setting of PD, has not received other anticancer treatments for their disease, and does not meet any of the IP discontinuation criteria as described in the protocol. Only 1 round of re-treatment will be allowed.

Nivolumab will be administered according to the package insert guidelines.

2.3 Treatment Assignment and Blinding

The study is not blinded.

During the dose-escalation phase, each subject who meets the eligibility criteria will be assigned open-label IP(s). A subject is considered entered into the study when the investigator notifies the IXRS that the subject meets eligibility criteria and is enrolled into the study.

During the dose-expansion phase, subjects will be randomized 2:1 to either receive MEDI0680 in combination with durvalumab, or nivolumab monotherapy. Subjects are considered randomized into the study when the site confirms eligibility and enters the subject into the IXRS and are assigned unblinded IP(s).

2.4 Sample Size

A total of approximately 96 subjects could be required for both the dose-escalation and dose-expansion phases. For dose-escalation, the number of subjects will depend upon the toxicities observed as the study progresses. Up to approximately 36 evaluable subjects could be required by the 3 + 3 design with 6 dose treatment groups in the main dose-escalation design. Additional subjects could be required if dose de-escalation occurs, intermediate

dosing using the modified zone-based design is explored, the dosing interval of MEDI0680 is changed, or expansion of a dose-escalation treatment groups occurs. Subjects in the dose-escalation phase are considered evaluable if they receive the protocol-assigned doses of both durvalumab and MEDI0680, and complete the DLT-evaluation period or experience a DLT during the DLT-evaluation period. Non-evaluable subjects will be replaced. The table below provides the probability of dose-escalation to the next higher level for each underlying true DLT rate. For example, for a common toxicity that occurs in 10% of subjects, there is a greater than 90% probability of escalating to the next higher dose level. Conversely, for a toxicity that occurs with a rate of 60%, the probability of escalating to the next higher dose level is less than 10%.

Table 2.1 Probability of Escalating Dose for Different True Underlying Dose-limiting Toxicity Rate at Given Dose Level

True Underlying DLT Rate	10%	20%	30%	40%	50%	60%	70%	80%	90%
Probability of escalating dose	0.91	0.71	0.49	0.31	0.17	0.08	0.03	0.009	0.001

DLT = dose-limiting toxicity.

For the dose-expansion phase, up to approximately 60 subjects (40 subjects in the MEDI0680/durvalumab combination therapy arm and 20 subjects in the nivolumab monotherapy arm) may be randomized at the selected combination dose (ie, MTD or highest protocol-defined dose for each agent in the absence of exceeding the MTD) in a 2:1 ratio. Assuming an ORR for nivolumab monotherapy of 21.5%, the sample size is chosen to detect a difference in ORR of 26.0% (ie, ORR = 47.5%) with 76% power at a 1-sided significance level of 0.10. The 95% CI of a 47.5% ORR (19 responders / 40 subjects) based on the exact probability method is 31.5%, 63.9%. The assumption of a nivolumab monotherapy ORR of 21.5% is based on the data from the nivolumab package insert.

3 STATISTICAL METHODS

3.1 General Considerations

Categorical data will be summarized by frequency distribution (number and percentage of subjects falling within each category). Continuous variables will be summarized by descriptive statistics including N, mean, standard deviation, median, and range (minimum

and maximum). Unless otherwise specified, all analyses will be based on the As-treated Population, which includes all subjects who receive any dose of IP analyzed according to the treatment they actually receive. All available data will be used and thus, missing data will not be imputed. In general, subjects with missing data for a parameter will be excluded from the summary of this parameter. Tables will be supported by data listings showing the original data forming the basis for the summaries. All data will be provided in data listings sorted by treatment group, subject number and date collected where applicable.

All subjects who are treated in the escalation phase will be summarized by treatment group.

All subjects in the expansion phase will be summarized by treatment groups

Two-sided confidence intervals, whenever specified, will be produced at 95%, unless otherwise specified.

In general, unless otherwise specified, Cycle 1/Day 1 (predose) will be the baseline assessment unless data are missing, in which case, baseline will be defined as the last value prior to dosing.

The data analyses will be conducted using the SAS® System (SAS Institute, Inc., Cary, NC, USA) Version 9.3 or above in Unix (Sun OS) environment. All analysis outputs will be validated according to MedImmune SAS® programming standards and MedImmune SAS® validation procedures.

3.2 Analysis Populations

The analysis populations are defined in Table 3.1.

Table 3.1 Analysis Populations

Population	Description
As-treated Population	Includes all subjects who receive any durvalumab (MEDI4736) or MEDI0680 (AMP-514) or nivolumab will be included in the As-treated Population and subjects will be analyzed according to the treatment they actually received. The As-treated Population will be used to evaluate baseline characteristics as well as all endpoints for the safety and efficacy profiles.

Table 3.1 Analysis Populations

Population	Description
Re-treated Population	Includes all PD subjects who are re-administered durvalumab (MEDI4736) or MEDI0680 (AMP-514) for up to 12 months after the initial study treatment period. The Re-treated Population will be used to evaluate safety and efficacy endpoints for re-treated subjects. The Re-Treated Population will be presented as separate summaries if enough subjects are included in the population.
DLT Evaluable Population	Includes all subjects enrolled in the dose-escalation phase who receive at least 2 protocol-assigned doses of durvalumab (MEDI4736) and MEDI0680 (AMP-514) for MEDI0680 (AMP-514) every 2-week dosing schedule and complete the safety follow-up through the DLT evaluation period (until the planned administration of the third dose for MEDI0680 (AMP-514) every 2-week dosing schedule or experience a DLT during the DLT evaluation period.
Response Evaluable Population	Includes any subject included in the As-treated Population who has a baseline disease assessment (DA) and either has at least one post-baseline DA and/or discontinued treatment due to death or disease progression prior to a post-baseline DA.

3.3 Study Subjects

3.3.1 Subject Disposition and Completion Status

Summaries of the number and percentage of subjects entered at each site will be provided.

Reasons for subjects discontinuing treatment will be summarized in terms of the number and percent of subjects.

The subject status at the end of study will be summarized in terms of the number and percent of subjects. For the subjects who end the study due to withdrawal of consent, the reason for withdrawal will be summarized.

The end of study mortality summary will include subjects who are dead at the end of study and their cause of death (toxicity related to IP, disease under investigation, or other).

3.3.2 Demographics and Baseline Characteristics

Baseline disease characteristics will be summarized for the As-treated Population. An additional summary may be presented for the Re-treated Population. Demographic information related to sex, age, race, ethnicity, weight, and height will be summarized. Tumor history including histology, stage, and pertinent biomarker results at the time of initial diagnosis and at study entry will be summarized.

The summary for prior cancer treatment will include the number and percent of subjects by treatment category (biologic, chemotherapy, growth factors, other), number of lines of previous systemic therapy, and best response (complete response, partial response, stable disease, progressive disease, not evaluable, not done) to the most recent line of therapy.

3.3.3 Investigational Product Exposure

Treatment exposure will be summarized for the As-treated Population. An additional summary may be presented for the Re-treated Population. Duration of exposure to durvalumab (MEDI4736), MEDI0680 (AMP-514) and nivolumab in months and cycles will be summarized by descriptive statistics and by frequency. Dose intensity and relative dose intensity of durvalumab (MEDI4736), MEDI0680 (AMP-514) and nivolumab will be summarized by descriptive statistics. The relative dose intensity for durvalumab (MEDI4736), MEDI0680 (AMP-514) and nivolumab is a percent of total actual dose that a subject received during corresponding study treatment period versus the total intended dose for the same study treatment period according to the study protocol. The details of the dose intensity calculation will be provided in the SPP.

Dosing delays, omissions, and interruptions of durvalumab (MEDI4736), MEDI0680 (AMP-514) and nivolumab will be summarized at the subject level and for total doses delayed, as well as reason for delay with frequency distributions.

The use of subsequent anticancer treatment after the discontinuation of study treatment will be summarized by the type of treatment.

3.3.4 Concomitant Medications

Concomitant medications will be summarized for the As-treated Population. An additional summary may be presented for the Re-treated Population. The number and percentage of subjects who received at least 1 dose of medication other than durvalumab (MEDI4736), MEDI0680 (AMP-514) or nivolumab during the study will be summarized by the generic name coded by Astrazeneca Drug Dictionary (AZDD) dictionary for each treatment group.

The summary table of concomitant medications will include all concomitant medications taken on or after the date of first dose of durvalumab (MEDI4736), MEDI0680 (AMP-514) or nivolumab or any concomitant medication started prior to first dose of study treatment that continued beyond the date of first dose of durvalumab (MEDI4736), MEDI0680 (AMP-514) or nivolumab .

3.4 Efficacy Analyses

3.4.1 Efficacy Endpoints and Analyses

The efficacy endpoints include best overall response (BOR); objective response (OR); disease control (DC); time to response (TTR); duration of response (DR); progression-free survival (PFS); overall survival (OS); and change from baseline in target lesion sum of diameters. All efficacy analyses will be analyzed based on the As-treated Population. Additional summaries may be presented for the Re-treated and/or Response-evaluable Populations. Time to disease assessments (TDA) will also be summarized.

For the dose escalation phase, efficacy analyses will be based on both an application of Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 to investigator assessed tumor measurements and an application of *modified* RECIST 1.1 to investigator assessed tumor measurements. For the latter, RECIST 1.1 has been modified to require confirmation of PD. A confirmed PD will be a PD confirmed by a consecutive repeat assessment no fewer than 4 weeks later. A PD that occurs without follow-up scans to provide confirmation or only non-evaluable follow-up scans will also be considered a confirmed PD. No such modification will be made in the dose expansion phase.

In the expansion phase, efficacy analyses will be based on an application of RECIST 1.1 to investigator assessed tumor measurements and may also be based on the Blinded Independent Central Review (BICR) if such data is collected.

Time-to-event data will be summarized using Kaplan-Meier estimates (median time and two-sided 95% CI for median time). An 80% CI for median time *may* also be included.

3.4.1.1 Best Overall Response

Best overall response (BOR) will be based on all post-baseline disease assessments that occur prior to the initiation of subsequent anticancer treatment. If re-treatment is permitted following an initial disease progression, a BOR will be reported separately for the re-treatment period. BOR will be summarized with the number and percentage of subjects for

the following categories: complete response (CR); partial response (PR); stable disease (SD); progressive disease (PD); and non-evaluable (NE).

At least 6 weeks from randomization (dose expansion) / the first dose of IP (dose escalation) first dose of IP must elapse without a radiological disease progression in order to assign a best overall response of SD. The definition of 6 weeks will allow for the protocol-defined disease assessment window of 3 days. Confirmation of PR and CR is required and must occur no fewer than 4 weeks after initial documentation of PR or CR. If a PR or CR that is pending confirmation is designated at an assessment followed by 1 or more non-evaluable assessments for CR or 1 or more non-evaluable assessments or assessments of SD for PR, response may be confirmed thereafter. Confirmation of PD is required in the dose escalation phase, but not the dose expansion phase. A confirmed PD will be a documented occurrence of a PD that is confirmed by a consecutive repeat assessment no fewer than 4 weeks later. A PD that occurs without follow-up scans to provide confirmation or only non-evaluable follow-up scans will also be considered a "confirmed" PD.

Concordance/discordance of disease assessment by investigator and derived disease assessment will be summarized in frequency table.

3.4.1.2 Objective Response

Objective response rate (ORR) is defined as the proportion of subjects with a BOR of confirmed CR or confirmed PR. ORR will be estimated with a 95% CI using the exact probability method. For the expansion phase, an estimate of the difference in ORR between treatment groups will be reported and tested for significance with a Pearson chi-square test (or Fisher's exact test, as appropriate). Subjects that have missing overall response assessments will be considered non-responders, and will therefore be counted in the denominator, but not in the numerator of ORR.

3.4.1.3 Disease Control

Disease control rate - 8 (DCR8) is defined as the proportion of subjects with a BOR of confirmed CR, confirmed PR, or SD (maintained for ≥ 8 weeks). A similar definition applies to DCR24. Duration of SD is defined as the time from randomization (dose expansion) / the first dose of IP (dose escalation) until the first documentation of disease progression or death due to any cause, whichever occurs first, and will allow for the protocol-defined disease assessment window of 3 days. DCR8 and DCR24 will be estimated with a 95% CI using the exact probability method. For the expansion phase, an estimate of the difference in DCR8 and DCR24 between treatment groups will be reported and tested for significance with a Pearson chi-square test (or Fisher's exact test, as appropriate).

3.4.1.4 Time to Response

Time to response (TTR) is defined as the time from randomization (dose expansion) / the first dose of IP (dose escalation) until the first documentation of objective response. Only subjects who have achieved objective response will be evaluated for TTR. TTR is defined in months as follows:

> TTR (months) = (Date of first confirmed disease response – Date of randomization (dose expansion) / the first dose of IP (dose escalation) first dose of IP + 1) / (365.25/12).

The median TTR and its 95% CI will be assessed using the Kaplan-Meier method. An 80% CI for median time *may* also be included.

3.4.1.5 Duration of Response

Duration of response (DR) is defined as the time from the first documentation of a subsequently confirmed objective response until the first documentation of disease progression or death due to any cause, whichever occurs first. Only subjects who have achieved objective response will be evaluated for DR. DR is defined in months as follows:

> DR (months) = (Date of PD/death or censoring – Date of first subsequently confirmed disease response + 1) / (365.25/12),

such that date of PD/death or censoring is given in Table 3.2. The median DR and its 95% CI will be estimated using the Kaplan-Meier method. An 80% CI for median time *may* also be included.

3.4.1.6 Progression-free Survival

Progression-free survival (PFS) is defined as the time from randomization (dose expansion) / the first dose of IP (dose escalation) until the first documentation of disease progression or death due to any cause, whichever occurs first, regardless of whether the subject receives subsequent anticancer therapy prior to progression. Subjects who have not progressed at the time of analysis will be censored at the time of the latest date of assessment from their last adequate progression free disease assessment. Censoring rules are presented in Table 3.2. PFS is defined in months as follows:

> PFS (months) = (Date of PD/death or censoring – Date of randomization (dose expansion) / the first dose of IP (dose escalation) first dose of IP+ 1) / (365.25/12)

The median PFS and its 95% CI will be estimated using the Kaplan-Meier method. An 80% CI for median time *may* also be included.

For the expansion phase, the difference in PFS between treatment groups will be tested for significance with a logrank test. The hazard ratio for the two treatment groups with two-sided 95% CI will be estimated by a Cox proportional hazard model controlling for the prespecified stratification factors as explanatory variables in the model. The proportion of subjects progression free and alive at 6 months and associated 95% CI will be estimated using the Kaplan-Meier method.

CENSORING OF PFS

Table 3.2 Summary of Censoring Guidelines for Efficacy Endpoints

Situation	Date of PD/Death or Censoring	PFS Outcome
Documented Progressive Disease (PD)	Date of earliest sign of PD	Event
Death prior to second scheduled post-baseline disease assessment or after an adequate post-baseline disease assessment	Date of death	Event
No PD or death at time of analysis or lost to follow-up	Date of last adequate progression-free disease assessment	Censored
Death or PD immediately after ≥ 2 consecutive missed or non-evaluable disease assessments	Date of randomization (dose expansion) / the first dose of IP (dose escalation) or last progression-free disease assessment prior to missed or non-evaluable assessments, whichever occurred last	Censored
Initiation of subsequent anticancer treatment	Date of last adequate progression-free disease assessment prior to initiation of subsequent anticancer treatment	Censored for Sensitivity Analyses
No tumor assessment at baseline and no evidence of PD at first post-baseline disease assessment, OR No tumor assessment post-first dose, and no death prior to second scheduled post-baseline disease assessment	Date of randomization (dose expansion) / the first dose of IP (dose escalation)	Censored

PD = progressive disease; PFS = progression-free survival

Subjects having missing lesion data at baseline or no disease assessments post-first dose will have PFS censored at the date of randomization (dose expansion) / the first dose of IP (dose

escalation) unless the subject dies prior to the second scheduled post-baseline disease assessment in which case the death date will qualify as a PFS event.

If a subject has two or more consecutive completely missed or non-evaluable assessments followed immediately by death or an assessment showing radiologic disease progression, then the subject will be censored for PFS. PFS will be censored at the date of randomization (dose expansion) / the first dose of IP (dose escalation) or the last progression-free disease assessment prior to the missed or non-evaluable assessments, whichever occurred last. If a subject has two or more consecutive missed or non-evaluable assessments followed by an assessment showing no radiologic disease progression, then the assumption will be that the subject did not progress during the missed or non-evaluable assessments. Two or more consecutive assessments is defined as ≥ 17 weeks (two disease assessments as per protocol plus a 1 week visit window to allow for a late assessment) after the last evaluable post-baseline disease assessment.

Subjects remaining on study without radiologic disease progression or death at the time of analysis will be censored for PFS at the date of their last adequate disease assessment. An adequate assessment is one that has recorded measurements for all target lesions and statuses for all non-target lesions defined at baseline (ie, the last non-missing or non-evaluable disease assessment). Additionally, a sensitivity analysis may be performed censoring PFS at the date of the last progression-free disease assessment prior to initiation of subsequent anticancer treatment.

3.4.1.7 Overall Survival

Overall survival (OS) is defined as the time from randomization (dose expansion) / the first dose of IP (dose escalation) until death due to any cause. A subject alive at the end of study or lost to follow-up will be censored for OS at the last date when the subject was known to be alive. OS is defined in months as follows:

> OS (months) = (Date of death or censoring – Date of randomization (dose expansion) / the first dose of IP (dose escalation) + 1) / (365.25/12).

The median OS and its 95% CI will be estimated using the Kaplan-Meier method. For the expansion phase, the difference in OS between treatment groups will be tested for significance with a logrank test. The hazard ratio for the two treatment groups with two-sided 95% CI will be estimated by a Cox proportional hazard model controlling for the prespecified stratification factors as explanatory variables in the model. The proportion of subjects alive at 12 months and associated 95% CI will be estimated using the Kaplan-Meier method.

3.4.1.8 Change from Baseline in Tumor Sizes

The percent change from baseline in target lesion sum of diameters (longest for non-nodal lesions, short axis for nodal lesions) will be calculated at each post-baseline disease assessment with recorded measurement for all target lesions defined at baseline. The percent change from baseline in target lesion sum of diameters is defined as follows:

100 * (Σ Diameters at DA X - Σ Diameters at BL) / (Σ Diameters at BL).

The percent change from baseline in target lesion sum of diameters will be presented by subject using spider plots. The best percent change from baseline in target lesion sum of diameters is defined as the largest reduction or smallest increase (in the case where a reduction does not occur) from baseline observed over all post-baseline disease assessments. The best percent change from baseline will be presented using waterfall plots. Target lesion measurements and sum of diameters will be listed by disease assessment and subject.

3.4.1.9 Time to Disease Assessments

Time to Disease Assessments (TDA) is defined as the time from randomization (dose expansion) / the first dose of IP (dose escalation) first dose of IP until each disease assessment, regardless of evaluability. TDA is defined in days as follows:

TDA (days) = (Date of disease assessment – Date of randomization (dose expansion) / the first dose of IP (dose escalation) first dose of IP+ 1).

The median TDA and its 95% CI will be estimated using the Kaplan-Meier method for the first 4 post-baseline disease assessments. For disease assessments that span multiple days, the disease assessment date assigned for the TDA analysis will be the date when the first imaging scan from the disease assessment was performed.

3.4.2 Additional Analyses of the Efficacy Endpoints

Disease Response will also be assessed based on an application of immune-related Response Evaluation Criteria in Solid Tumors (irRECIST) (Nishino et al, 2013) to investigator assessed tumor measurements. All new lesions meeting measureable criteria will be included in the Tumor Burden. Objective Response by irRECIST is defined as immune-related complete response (irCR) or immune-related partial response (irPR) with confirmation required occurring no fewer than 4 weeks after initial documentation of irCR or irPR. If an irPR or irCR that is pending confirmation is designated at an assessment followed by 1 or more non-evaluable assessments or assessments of irSD, response may be confirmed thereafter Disease Control - 8 by irRECIST is defined as confirmed irCR, confirmed irPR, or immune-related stable disease (irSD) (maintained for ≥ 8 weeks). A similar definition applies

to Disease Control - 24. Duration of irSD is defined as the time from randomization (dose expansion) / the first dose of IP (dose escalation) until the first documentation of disease progression or death due to any cause, whichever occurs first, and will allow for the protocol-defined disease assessment window of 3 days. Concordance/discordance between RECIST 1.1 and irRECIST disease assessment will be summarized in frequency table.

3.4.3 Subgroup Analyses

No subgroup analyses are planned.

3.3.5 Response Derivations and Handling of Missing Data

Guidance regarding the handling of dropouts and missing data will apply uniformly to all efficacy analyses resulting from an application of RECIST 1.1 and irRECIST to investigator assessed tumor measurements. For investigator reported outcomes, analyses will present outcomes reported by the investigator without consideration of missing data or censoring rules.

3.3.5.1 RECIST 1.1 Derivations

DERIVATION OF TARGET LESION RESPONSE

Target lesion response will be programmatically derived on the data collection instrument once RECIST 1.1 criteria are applied to the site personnel recorded target lesion measurements.
Possible values include:

- CR – Complete Response
- PR – Partial Response
- SD – Stable Disease
- PD – Progressive Disease
- NE – Non-evaluable
- NA – Not Applicable (*set value for all post-baseline disease assessments only if no target lesions are identified at baseline*)

The derivation for target lesion response is as follows *(please note the order of the algorithm below is important)*:

1. If "Any Target Lesions Present" equals "No" on the *Target Lesions – Baseline* CRF, then all post-baseline "Target Lesion Response" equals "NA".
2. Else, if "Percent Change from Nadir Sum of Diameters" is greater than or equal to 20% and the absolute increase from the nadir (defined as the "Total" for each post-

baseline disease assessment minus the "Nadir Sum of Diameters") is greater than or equal to 5 mm, then "Target Lesion Response" equals "PD".

3. Else, if "Not Done" is selected, or "Measurement" is left blank, or "Lesion no longer Measureable" is selected and equal to "NE", or "Lesion Intervention" is selected for any Target Lesion identified at Baseline, then "Target Lesion Response" equals "NE".
4. Else, if "Total Non-Lymph Node" equals "0" and all Lymph Node Target Lesion "Measurements" are less than "10" individually, then "Target Lesion Response" equals "CR".
    a. Note: This step requires examining "Measurements" separately for Target Lesions with "Lymph Node" equal to "Yes" and "No".
5. Else, if "Percent Change from Baseline Sum of Diameters" is less than or equal to -30%, then "Target Lesion Response" equals "PR".
6. Else, "Target Lesion Response" equals "SD".

If a subject has a missing tumor measurement at a disease assessment for 1 or more target lesions, the sum of diameters (longest for non-nodal lesions, short axis for nodal lesions) will be reported for the remaining target lesions. These data will be used to indicate radiologic disease progression if the sum of diameters for the observed lesions increases at least 20% from the nadir sum of diameters of all lesions and demonstrates at least a 5 mm absolute increase from the nadir sum of diameters of all lesions, in spite of the missing data (or if other criteria for PD are met).

DETERMINATION OF NON-TARGET LESION RESPONSE

Non-target lesion response will be assigned by site personnel following a qualitative overall assessment of all non-target lesions.
Possible values include:

- CR – Complete Response
    - Disappearance of all non-target lesions and normalization of tumor marker level. All lymph nodes must be non-pathological in size (<10mm short axis)
- Non-CR/Non-PD – Non-Complete Response / Non-Progressive Disease
    - Persistence of one or more non-target lesion(s) and/or maintenance of tumor marker level above the normal limits
- PD – Progressive Disease
    - Unequivocal progression of existing non-target lesions.
- NE – Non-evaluable
- NA – Not Applicable (*set value for all post-baseline disease assessments only if no non-target lesions are identified at baseline*)

Though non-target lesion responses are a subjective decision made by the site personnel, certain responses may be limited depending on the non-target lesion statuses recorded. An algorithm is provided below highlighting appropriate possible non-target lesion responses based on recorded data. Reaching a red box (■) signifies having reached the only allowable non-target lesion responses based on non-target lesion statuses. Reaching a green box (■) signifies having reached the end of the algorithm and more than one possible non-target lesion response is possible from which the Investigator may choose.

1. a) If no non-target lesions are identified at baseline, all post-baseline non-target lesion responses should equal NA. ■
   b) Else, if any non-target lesions are identified at baseline, responses may be limited to CR, Non-CR/Non-PD, PD, NE (ie, responses of NA are not permitted).
   Go to Rule 2.
2. a) If all non-target lesions have a status are "Absent", the responses may be limited to CR. ■
   b) Else, if at least one non-target lesion status is NOT "Absent", the responses may be limited to Non-CR/Non-PD, PD, NE (ie, responses of CR, NA are not permitted).
   Go to Rule 3.
3. a) If all non-target lesions have a status of "Unequivocal Progression", responses may be limited to PD. ■
   b) Else, if no non-target lesions have a status of "Unequivocal Progression", responses may be limited to Non-CR/Non-PD, NE (ie, responses of CR, PD, NA are not permitted).
   Go to Rule 4.
   c) Else, if at least one (but not all) non-target lesion has a status of "Unequivocal Progression", the responses may be limited to Non-CR/Non-PD, PD, NE (ie, responses of CR, NA are not permitted). (*Note: No response has been eliminated as an option here.*)
   Go to Rule 5.
4. a) If all non-target lesions have a status of "Non-evaluable" and/or "Not Done" is selected, responses may be limited to NE. ■
   b) Else, if no non-target lesions have a status of "Non-evaluable" and "Not Done" is not selected, responses may be limited to Non-CR/Non-PD (ie, responses of CR, PD, NE, NA are not permitted). ■
   c) Else, if at least one (but not all) non-target lesion has a status of "Non-evaluable" and/or "Not Done" is selected, the responses may be limited to Non-CR/Non-PD, NE

(ie, responses of CR, PD, NA are not permitted). ■

(*Note: No response has been eliminated as an option here.*)

5. a) If all non-target lesions have a status of "Non-evaluable" and/or "Not Done" is selected, responses may be limited to NE. ■

   b) Else, if no non-target lesions have a status of "Non-evaluable" and "Not Done" is not selected, responses may be limited to Non-CR/Non-PD, PD (ie, responses of CR, NE, NA are not permitted). ■

   c) Else, if at least one (but not all) non-target lesion has a status of "Non-evaluable" and/or "Not Done" is selected, the responses may be limited to Non-CR/Non-PD, PD, NE (ie, responses of CR, NA are not permitted). ■

   (*Note: No response has been eliminated as an option here.*)

If a subject has a missing tumor status at a disease assessment for 1 or more non-target lesions, radiologic disease progression will be determined if the remaining non-target lesions qualitatively demonstrate unequivocal progression (or if other criteria for PD are met).

DERIVATION OF DISEASE RESPONSE AS PER RECIST 1.1

Disease response will be programmatically derived on the data collection instrument using RECIST 1.1 criteria based upon target lesion response, non-target lesion response, and new lesion data. Missing values in any of target lesion response, non-target lesion response, and new lesion data will result in the disease response not being derived.
Possible values include:

- CR – Complete Response
- PR – Partial Response
- SD – Stable Disease
- PD – Progressive Disease
- NE – Non-evaluable

Target Lesion Response	Non-Target Lesion Response	New Lesion	Derived RECIST Disease Response
CR	CR or NA	No	CR
CR	Non-CR/Non-PD or NE	No (or NE)	PR
PR	CR or Non-CR/Non-PD or NE or NA	No (or NE)	PR
SD	CR or Non-CR/Non-PD or NE or NA	No (or NE)	SD
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NE	CR or Non-CR/Non-PD or NE or NA	No	NE
NA	CR	No	CR
NA	Non-CR/Non-PD	No	SD (Non-CR/Non-PD)[a]
NA	NE or NA	No (or NE)	NE
NA	CR or Non-CR/Non-PD	NE	SD (Non-CR/Non-PD)[a]

[a] Per RECIST 1.1, "SD (Non-CR/Non-PD)" is preferred over "SD" for non-target disease since SD is increasingly used as endpoint for assessment of efficacy in some trials so to assign this category when no lesions can be measured is not advised.

If a subject has a missing tumor measurement at some assessment(s) for 1 or more target or a missing tumor status at some assessment(s) for 1 or more non-target lesions and criteria for PD are not otherwise met, an overall response of NE will be assigned for the assessment(s).

3.3.5.2 irRECIST Derivations

DERIVATION OF TUMOR BURDEN RESPONSE

Tumor burden will be programmatically derived on the data collection instrument once irRECIST criteria are applied to the site personnel recorded target lesion and new lesions included in tumor burden measurements.
Possible values include:

- irCR – Complete Response
- irPR – Partial Response
- irSD – Stable Disease
- irPD – Progressive Disease
- irNE – Non-evaluable
- irNA – Not Applicable (*set value for all post-baseline disease assessments only if no target lesions are identified at baseline*)

The derivation for tumor burden response is as follows *(please note the order of the algorithm below is important)*:

1. If "Any Target Lesions Present" equals "No" on the *Target Lesions – Baseline* CRF, then all post-baseline "Tumor Burden Response" equals "irNA".
2. Else, if "Percent Change from Nadir Sum of Diameters" is greater than or equal to 20% and the absolute increase from the nadir (defined as the "Total" for each Post-Baseline DA minus the "Nadir Sum of Diameters") is greater than or equal to 5 mm, then "Tumor Burden Response" equals "irPD".
3. Else, if "Not Done" is selected, or "Measurement" is left blank, or "Lesion no longer Measureable" is selected and equal to "NE", or "Lesion Intervention" is selected for any target lesion identified at baseline or new lesions included in the tumor burden, then "Tumor Burden Response" equals "irNE".
4. Else, if "Total Non-Lymph Node" equals "0" and all lymph node target lesion and new lesion included in the tumor burden "Measurements" are less than "10" individually, then "Tumor Burden Response" equals "irCR".

a. Note: This step requires examining "Measurements" separately for Target lesions and new lesion included in the tumor burden with "Lymph Node" equal to "Yes" and "No".

5. Else, if "Percent Change from Baseline Sum of Diameters" is less than or equal to -30%, then "Tumor Burden Response" equals "irPR".
6. Else, "Tumor Burden Response" equals "irSD".

If a subject has a missing tumor measurement for lesions included in tumor burden at a disease assessment for 1 or more target lesions and/or new lesions included in tumor burden, the sum of diameters (longest for non-nodal lesions, short axis for nodal lesions) will be reported for the remaining target lesions and new lesions included in tumor burden. These data will be used to indicate radiologic disease progression if the sum of diameters for the observed lesions increases at least 20% from the nadir sum of diameters of all lesions and demonstrates at least a 5 mm absolute increase from the nadir sum of diameters of all lesions, in spite of the missing data (or if other criteria for irPD are met).

DERIVATION OF NON-TUMOR BURDEN RESPONSE

Non-tumor burden will be programmatically derived on the data collection instrument once irRECIST criteria are applied to the site personnel recorded non-target lesion and new lesions not included in tumor burden.
Possible values include:

- Absent – Response assigned if all non-target lesions and new lesions not included in the tumor burden statuses equal Absent
- Present / Indeterminate – Response assigned if any of the non-target lesions or new lesions not included in the tumor burden statuses do not equal Absent

If a subject has a missing tumor status at a disease assessment for 1 or more non-target lesions and/or new lesions not included in tumor burden, radiologic disease progression will be determined if the remaining non-target lesions and new lesions not included in tumor burden qualitatively demonstrate unequivocal progression (or if other criteria for irPD are met).

DERIVATION OF DISEASE RESPONSE AS PER IRRECIST

Disease response will be programmatically derived on the data collection instrument using irRECIST criteria based upon tumor burden response and non- tumor burden response data. Missing values in any of tumor burden response or non-tumor burden response data will result in the disease response not being derived.
Possible values include:

- irCR – Complete Response
- irPR – Partial Response
- irSD – Stable Disease
- irPD – Progressive Disease
- irNE – Non-evaluable
- Abs – Absent
- PI – Present / Indeterminate

Tumor Burden Response	Non-Tumor Burden Response	Derived irRECIST Disease Response
irCR	Absent	irCR
irCR	Present / Indeterminate	irPR
irPR	Any	irPR
irSD	Any	irSD
irPD	Any	irPD
irNE	Any	irNE
irNA	Absent	Absent
irNA	Present / Indeterminate	Present / Indeterminate

Note: Non-tumor burden response only factors into the derived irRECIST disease response in the case of irCR vs irPR.

If a subject has a missing tumor measurement at some assessment(s) for 1 or more target lesions and/or new lesions included in tumor burden or a missing tumor status at some assessment(s) for 1 or more non-target lesions and/or new lesions not included in tumor burden and criteria for irPD are not otherwise met, an overall response of irNE will be assigned for the assessment(s).

3.3.5.3 Locoregional therapy

Any subject receiving locoregional therapy, including surgery, while on study that directly affects one or more of the target lesions selected at baseline and/or new lesions included in tumor burden will be identified. A subject with a subsequent response or SD/irSD will be considered to be non-evaluable at all disease assessments that occur on or after the date of locoregional therapy. Otherwise, the subject will be assessed ignoring the locoregional therapy.

3.3.5.4 Assignment of Dates of Disease Progression or Disease Response

For all analyses of endpoints resulting from an application of modified RECIST 1.1 (dose escalation phase), RECIST 1.1 (expansion phase) to investigator assessed tumor measurements, and irRECIST, there may be cases in which disease assessments span a series of dates. For establishing the start date of a subsequently confirmed response in which the disease assessment spans multiple days, the response date assigned will be the latest date of evaluations corresponding to the disease assessment. The date of latest assessment will also be assigned for a mid-study assessment showing SD as the date assigned for the purposes of censoring of duration of response, TTP and PFS.
The date of PD will be the first date at which any objective diagnostic test provides data indicating PD. Specifically, the date of PD will be the earliest of the following 3 dates:

- Date of PD as indicated by target lesions: If PD is triggered by a change in sum of diameters of target lesions, and the dates of evaluation of the target lesions vary for the same assessment, assign the first evaluation date among target lesions.
- Date of PD as indicated by non-target lesions: If the dates of evaluation of the non-target lesions vary for the same assessment, assign the first evaluation date for which any non-target lesion exhibits a status of Unequivocal Progression.
- Date of PD as indicated by new lesions: If multiple new lesions are identified and the dates of evaluation for the new lesions vary for the same assessment, assign the first evaluation date for which any new lesion is detected.

3.5 Patient Reported Outcomes

All statistics listed in this section will be summarized for the As-treated Population. An additional summary may be presented for the Re-treated Population.

3.5.1 Pain Questionnaire

Pain scores were planned as per protocol early in this trial but were removed during the dose escalation phase in Protocol Amendment 3. Pain scores will be listed.

3.6 Pharmacodynamic Endpoints and Analyses

3.6.1.1 Pharmacodynamic Endpoints

Exploratory pharmacodynamic endpoints are listed in the protocol.

3.6.1.2 Analysis of Pharmacodynamic Endpoints

The pharmacodynamic endpoints will be analyzed by the MedImmune Translational Sciences group or designee.

3.7 Other Additional Analyses

Not applicable.

3.8 Safety Analyses

All statistics listed in this section will be summarized for the As-treated Population. An additional summary may be presented for the Re-treated Population. If the number of subjects in the Re-treated Population with available data is less than 5% of the total number of subjects in the As-treated Population for overall summary (across disease types) or < 5 subjects for one disease type at one scheduled time of evaluation, no summary statistics will be presented at that time point for overall summary or for that disease type, unless otherwise indicated.

3.8.1 Maximum Tolerated Dose

The Maximum Tolerated Dose (MTD) will be based on the DLT Evaluable Population. In summary, the number and percentage of subjects with a DLT during the Dose Escalation Phase will be presented by treatment groups. The MTD level will be indicated in the summary.

3.8.2 Adverse Events and Serious Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as events present at baseline that worsen in intensity after administration of IP or events absent at baseline that emerge after administration of IP. IP TEAEs will be coded by MedDRA version 17.1 or more recent version and the type incidence, severity graded according to NCI CTCAE version 4.03, and relationship to IP will be summarized. Specific adverse events will be counted once for each subject for calculating percentages. In addition, if the same adverse event occurs multiple times within a particular subject, the highest severity and level of relationship observed will be reported. If any associations of interest between adverse events and baseline characteristics are observed, additional stratified results may be presented. All treatment-emergent adverse events will be summarized overall, as well as categorized by MedDRA System Organ Class and Preferred Term.

The AEs and SAEs occurring from the signing of the informed consent and prior to the initiation of study treatment will be listed.

3.8.3 Adverse Events of Special Interest

An adverse event of special interest (AESI) is one of scientific and medical interest specific to understanding of the IP and may require close monitoring and rapid communication by the investigator to the sponsor. A list of AESIs for this study can be found in the protocol. Other categories may be added or existing terms may be modified as necessary.

3.8.4 Deaths and Treatment Discontinuations due to Adverse Events

For those subjects who died due to an AE, the AE contributing to death and the AE's relationship to IP will be summarized.

Summaries will be provided for TEAEs resulting in permanent discontinuation of IP. Supporting listings will be provided for AEs resulting in death and AEs resulting in permanent discontinuation of IP.

3.8.5 Clinical Laboratory Evaluation

Laboratory tests will be grouped according to chemistry, hematology, and urinalysis. Listings will be provided for all laboratory results, including urinalysis. For chemistry and hematology tests, the change in each laboratory parameter from baseline to each post-baseline visit will be summarized graphically.

For chemistry and hematology tests, laboratory abnormalities with toxicity grades according to the NCI CTCAE will be derived according to laboratory values. Laboratory abnormalities occurring from the start of IP administration to the last assessment on study will be presented. Tables will be generated showing the maximum toxicity grade, as well as any shifts of more than 2 grades from baseline to the maximum grade and to the last assessment on-study.

3.8.5.1 Hepatic Function Abnormality (Hy's Law)

Hepatic function abnormality meeting the definition of Hy's law (ie, any increase in ALT or AST to greater than 3 × ULN and concurrent increase in total bilirubin to be greater than 2 × ULN) is considered an AESI and will be reported. Concurrent findings are those that derive from a single blood draw or from separate blood draws taken within 8 days of each other. Follow-up investigations and inquiries will be initiated promptly by the investigational site to determine whether the findings are reproducible and/or whether there is objective evidence that clearly supports causation by a disease (eg, cholelithiasis and bile duct obstruction with distended gallbladder) or an agent other than the investigational product.

3.8.6 Other Safety Evaluations

3.8.6.1 Vital Signs

Vital signs will be assessed at baseline and throughout the study. Vital signs will be summarized graphically to illustrate changes from baseline.

3.8.6.2 Electrocardiogram (ECG)

Electrocardiogram (ECG) parameters will be assessed at baseline as well as throughout the study. Absolute values and maximum changes from baseline to post-therapy in QTc will be summarized using descriptive statistics.

3.8.6.3 ECOG Performance Status

Eastern Cooperative Oncology Group (ECOG) performance status will be assessed at baseline as well as throughout the study. ECOG will be summarized by study visit and will include descriptive statistics for the value of the parameters and the changes from baseline.

3.9 Immunogenicity

Analyses to assess the immunogenic potential of durvalumab (MEDI4736) will be performed by the MedImmune Global Pharmacokinetics - Pharmacodynamics (PK-PD) & Bioanalysis group or designee.

3.10 Pharmacokinetics

Pharmacokinetic data analyses will be performed by the MedImmune Global PK-PD & Bioanalysis group or designee.

4 INTERIM ANALYSIS

Interim futility analyses for MEDI0680/durvalumab combination vs nivolumab will be performed in a continuous manner using Bayesian predictive probability. Bayesian predictive probability is used in the setting of single-arm trials with tumor response endpoints such as OR and allows for continuous assessments for early Go/NoGo decision making (Lee and Liu, 2008). A joint Bayesian predictive probability approach has been developed which allows for continuous assessments of the delta (δ), or difference, of the ORRs between MEDI0680/durvalumab combination and nivolumab.

The interim analyses will begin after 20 subjects in the MEDI0680/durvalumab combination have been randomized and have reached their second post-baseline disease assessment or have completed study, whichever occurs first. Subject accrual may be paused in both treatment groups for the first interim analyses; however, an accrual pause may not be necessary if existing data upon enrollment of 20 subjects in the MEDI0680/durvalumab combination suggest futility criteria have not been triggered as per the predefined futility stopping criteria defined below. If futility is not triggered, further enrollment will continue without pause and continuous monitoring for futility will commence after an additional 4 subjects reach their second post-baseline disease assessment or completes the study, whichever occurs first, until enrollment is complete.

Given the existing observed data during the continuous monitoring stage, the joint predictive probability is obtained by calculating the probability of reaching a NoGo decision should the treatment groups be enrolled and evaluated to the maximum planned final sample size of 60 (40 in MEDI0680/durvalumab combination and 20 in nivolumab). Further enrollment into both treatment groups will be terminated due to futility if there is a high probability of reaching a NoGo decision upon enrollment of 60 subjects given the existing observed data (ie, predictive probability of a NoGo decision > 90%). An efficacy decision would be

reached if there is a high probability of reaching a Go decision upon enrollment of 60 subjects given the existing observed data (ie, predictive probability of a Go decision > 90%). If an efficacy decision is reached, enrollment may continue, but planning activities for further development of the MEDI0680/durvalumab combination may be triggered.

Decision matrices for continuous monitoring of ORR δ at the prespecified MEDI0680/durvalumab combination sample sizes for interim futility assessments are presented in Figure 4-1. The matrices presented are examples and assume that the MEDI0680/durvalumab combination treatment group will randomize in a 2:1 ratio to the nivolumab treatment group. Additional decision matrices will be prepared as needed to accommodate observed data. Futility, continuation, and efficacy decisions are marked using red, yellow, and green cells, respectively, and allow for an evaluation of varying combinations of responses between the two treatment groups. The target δ was set so as to demonstrate a 20% increase in the MEDI0680/durvalumab combination ORR over the benchmark nivolumab ORR based on investigator assessments, in a similar RCC population (ORR = 25.1 [95% CI: 21.0%, 29.6%]; Motzer et al, 2015). Given assumptions of a benchmark ORR in the nivolumab treatment group of 25% and a true δ = 20%, darker color cells represent scenarios that would be observed in practice with a 95% likelihood.

For example, given a target δ of 20% and assuming 3 observed responders in the nivolumab treatment group after 10 subjects have reached their second post-baseline disease assessment or have completed the study, observing 4 or fewer responders after 20 subjects in the MEDI0680/durvalumab combination have reached their second post-baseline disease assessment or have completed the study would trigger stopping enrollment claiming futility due to a high probability of reaching a NoGo decision upon enrollment to 60 subjects given the existing observed data Figure 4-2flo and Table 4.1).

A false negative is defined as observing either futility criteria during the continuous monitoring stage or NoGo criteria upon enrollment of 60 subjects in a given treatment group when the true ORR δ is greater than or equal to the target δ (ie, False Negative: P[NoGo Decision | True δ ≥ Target]). Assuming a true δ of 20%, the false negative rate is 11% (see highlighted cell in Table 4.1 below). Assuming a true δ of 25%, the false negative rate is 5%. Additionally, assuming a true δ of 20%, the Go decision rate (efficacy decision during the continuous monitoring stage or a Go decision upon enrollment of 60 subjects) is 37% with confirmation at the final analysis 75%.

Upon enrollment of 60 subjects in the dose-expansion phase, a final analysis will be performed. NoGo criteria will be met if the probability that the δ exceeds the pre-specified

target of 20% is less than 10% (ie, NoGo: P[δ > Target] < 10%). Go criteria will be met if the probability that the δ exceeds the pre-specified target of 20% is more than 70% (ie, Go: P[δ > Target] > 70%). For futility analyses, a response is defined as either a confirmed or unconfirmed CR or PR as per RECIST 1.1 and for the final analysis a response is defined as a confirmed CR or PR as per RECIST 1.1.

At interim analyses, efficacy and selected safety data will be summarized to assess the overall benefit-risk profile. The analysis populations to be used for the interim analysis are described below:

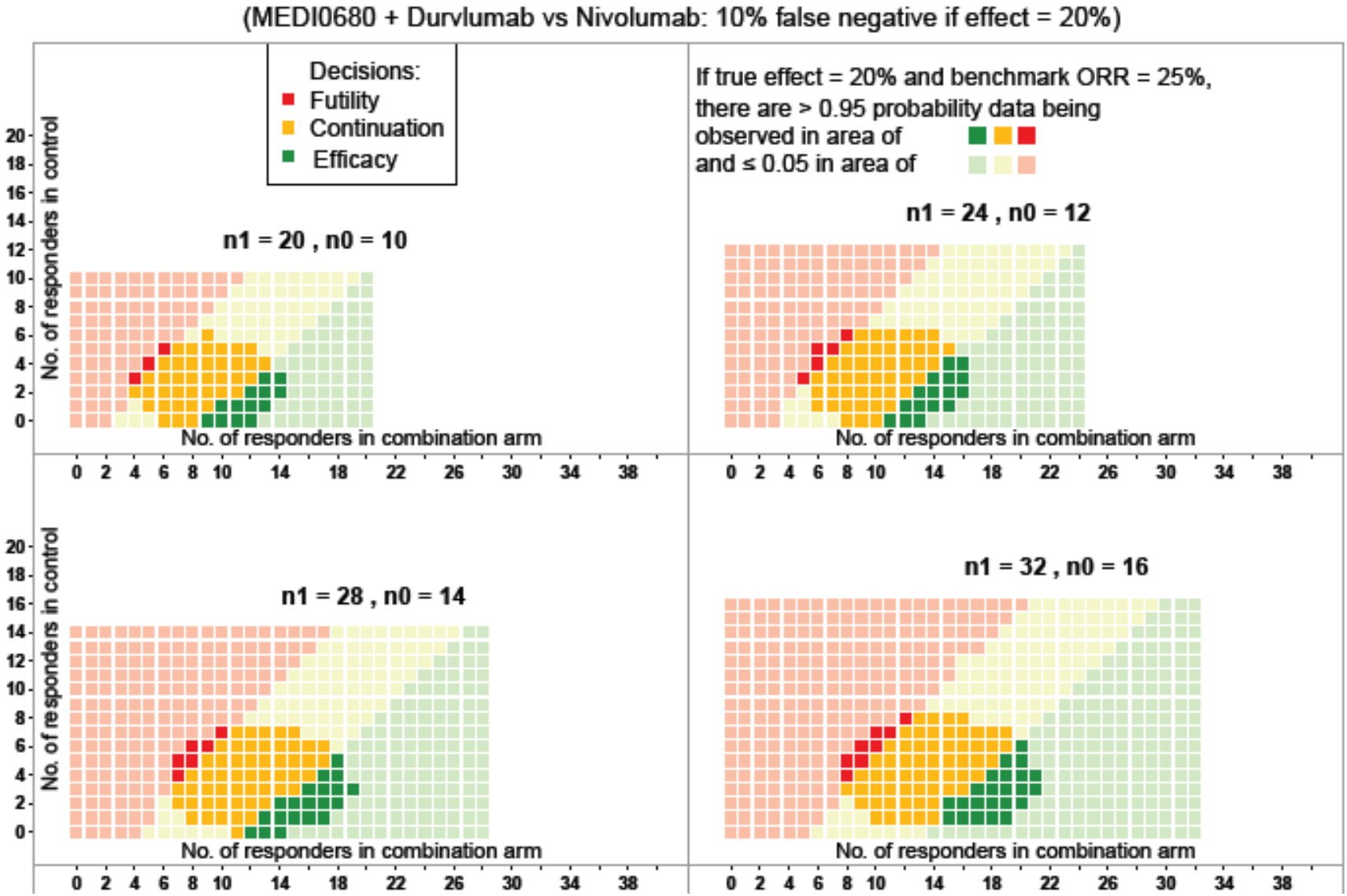
- **As-treated with 17 weeks of potential follow up** is defined as all subjects in the As-treated Population who have a baseline disease assessment and have an opportunity to be followed for at least 17 weeks (i.e., received the first dose of MEDI0680, durvalumab or nivolumab, whichever earliest, at least 17 weeks ago). The 17 weeks are to have at least 2 post-baseline disease assessments plus the one week visit window. This is the primary analysis population for the interim analysis of efficacy data.

- **Response evaluable with 17 weeks of potential follow up** is defined as all subjects in the As-treated Population who 1) have a baseline disease assessment; 2) have an opportunity to be followed for at least 17 weeks; and 3) have any post-baseline disease assessments or experience clinical PD/death. This is the secondary analysis population for the interim analysis of efficacy data.

- **As-treated with 9 weeks of potential follow up** is defined as all subjects in the As-treated Population who have a baseline disease assessment and have had an opportunity to be followed for at least 9 weeks. The 9 weeks are to have at least 1 post-baseline disease assessments plus one week visit window. This is the exploratory analysis population for the interim analysis of efficacy data.

- **As-treated Population**, as defined in section 3.2, is the primary analysis population for the interim analysis of safety data.

Figure 4-1 Decision Matrices for Continuous Monitoring



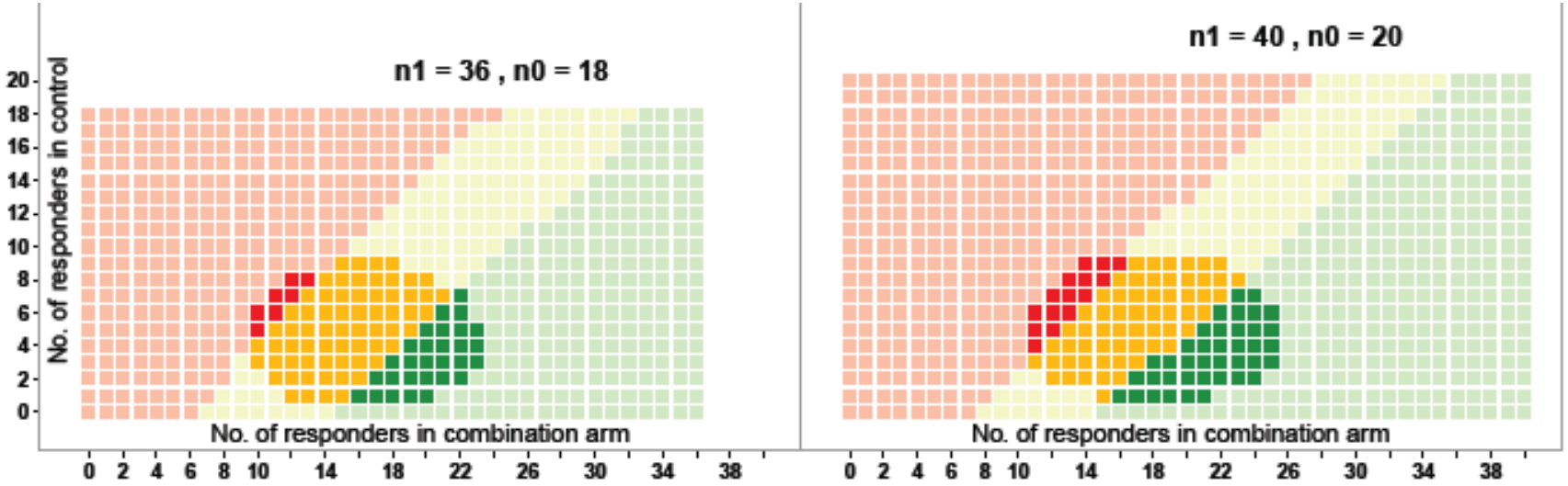
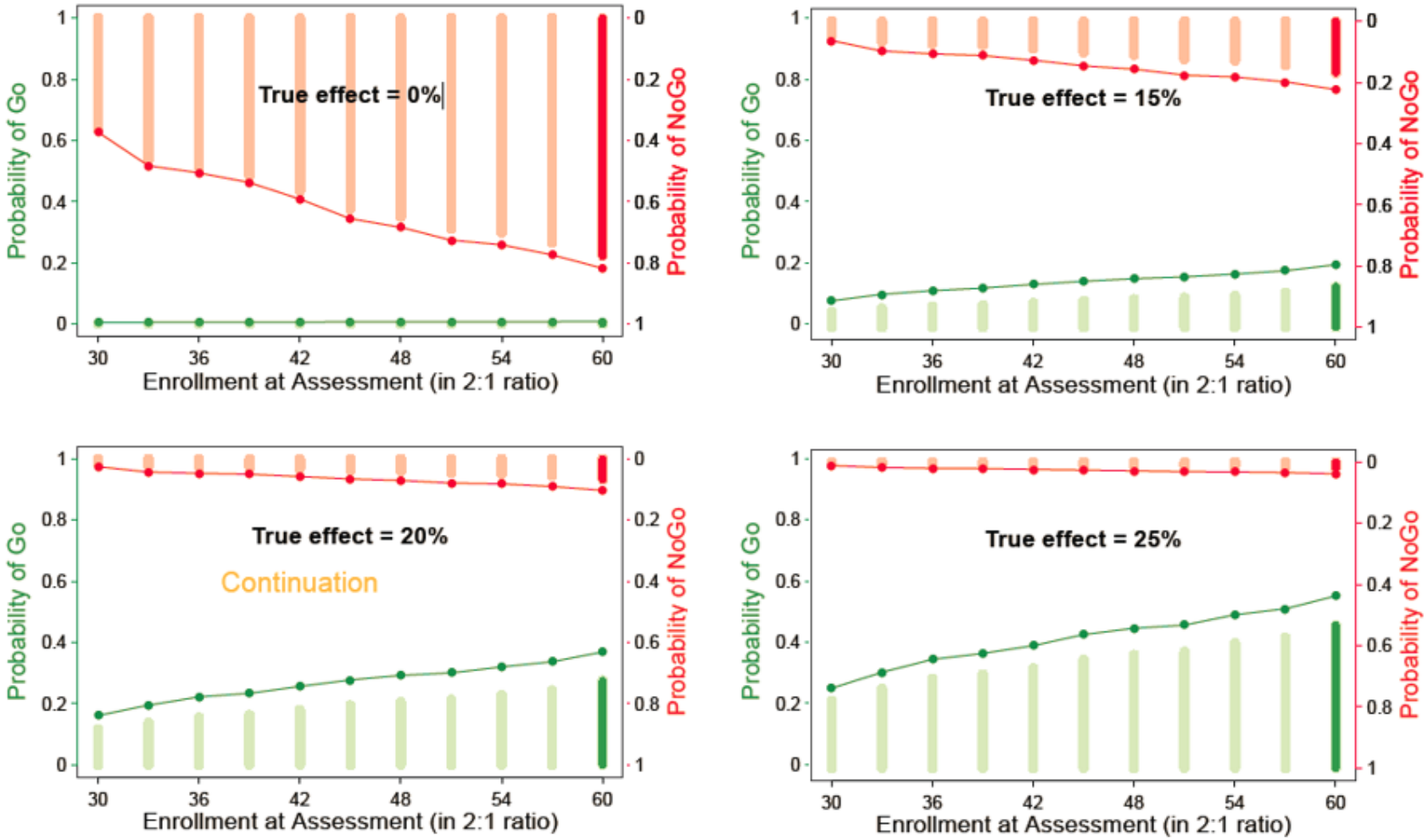

n1 = 36 , n0 = 18
n1 = 40 , n0 = 20
No. of responders in control
No. of responders in combination arm
No. of responders in combination arm
0 2 4 6 8 10 14 18 22 26 30 34 38
0 2 4 6 8 10 14 18 22 26 30 34 38
0 2 4 6 8 10 12 14 16 18 20

Figure 4-2 Visual Joint Predictive Probability Operating Characteristics



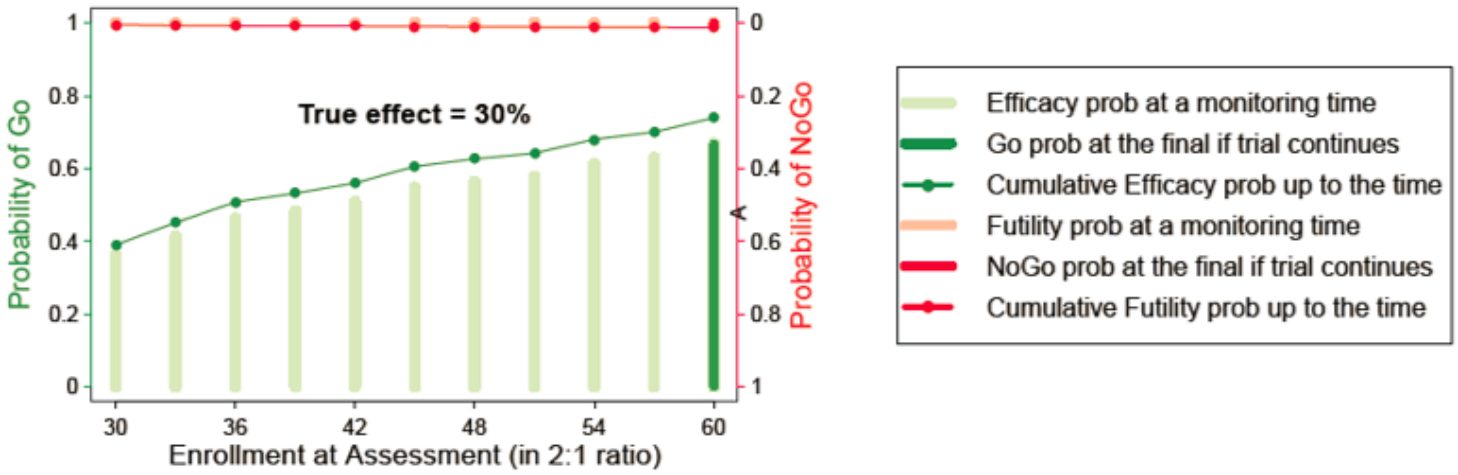


Table 4.1 Joint Predictive Probability Operating Characteristics

True Effect Size (Benchmark = 25%)	Decision based on Final Assessment				Decision based on Continuous Assessments			
	Go	NoGo	Consider		Efficacy (% confirmed with the final)	Futility	Expected Sample Size *up to a decision being made (% decision making prior to final)*	Expected Sample Size *stopping for futility only (% interim stopping)*
0%	1%	77%	23%		1% (28%)	82%	40 (78%)	41 (77%)
15%	13%	17%	70%		20% (65%)	24%	51 (39%)	55 (20%)
20%	28%	7%	65%		37% (75%)	11%	50 (43%)	58 (9%)
25%	47%	2%	51%		56% (84%)	5%	47 (56%)	59 (3%)
30%	67%	1%	32%		74% (90%)	2%	43 (72%)	60 (1%)

Abbreviations: PP: Predictive Probability. Target $\delta = 20\%$; Go: $Pr(\delta > 20\%) > 70\%$; NoGo: $Pr(\delta > 20\%) < 10\%$; Efficacy: $PP(Go) > 90\%$); Futility: $PP(NoGo) > 90\%$).

5 REFERENCES

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009; 45:228-47.

Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clin Trials. 2008; 5:93-106.

Motzer R, Escudier B, McDermott D, George S, Hammers H, Srinivas S, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. N Engl J Med. 2015; 373(19):1803-13.

Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. Clin Cancer Res. 2013; 19(4):3936-43.

7 VERSION HISTORY

Version	Date	Summary of Changes	Reason for Change
1.0	23Jun2014	Initial document	Initial document
2.0	03Nov2017	• The majority of the changes made to this amended SAP was to address the significant modifications to the study design made in Protocol Amendment 5. In particular, the primary update includes changing the standard of care treatment group from MEDI0680 Monotherapy to nivolumab in the Dose Expansion phase. Sections directly affected include: o Protocol Title o Section 2: Study Objectives / Study Design / Treatment Assignment / Sample Size o Section 3.4: Efficacy Analyses	• Protocol Amendment 5 resulted in several changes to the planned Dose Expansion phase.
		• Section 3 (Statistical Considerations): The entirety of this Section has been updated to incorporate current Standard SAP Statistical Consideration language. Content remains similar to previous SAP version.	• Modifications made to better fit with the more simplified and approved Standard SAP language
		• Section 4 (Interim Analysis): Interim futility analyses have been included for the Dose Expansion phase. The methods included in the SAP differ significantly from the low-bar version included in Protocol Amendment 5. The methods incorporated are the modified joint predictive probability which is an expansion to the single-arm approach to continuous monitoring that allows for the comparison of two treatment groups.	• Protocol Amendment 5 resulted in an addition of an Interim Analyses using continuous monitoring using predictive probability. Specifically, methods are used to incorporate data from active and SOC treatment groups.
3.0	06Mar2018	The change to this amended SAP is to add the analysis populations for the interim analysis.	To specify the analysis populations and corresponding definition for the interim analysis.