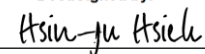MedImmune

MEMORANDUM

DATE: May 15, 2019

TO: Study File

FROM: Charles H. Darby

RE: D4190C00021 Statistical Analysis Plan Approval

The Statistical Analysis Plan (version 3.0) for Protocol D4190C00021 has been reviewed and approved.

DocuSigned by:

Teja Boppana

581F9FFB7E07488...

22 May 2019

Teja Boppana, Statistical Programming

DocuSigned by:

Charles H Darby

7D24622D7DEC4F6...

21 May 2019

Charles H Darby, Statistician

DocuSigned by:

Shahram Rahimian

42C1D7C6365B4D5...

21 May 2019

Shahram Rahimian, Clinical Development Lead

DocuSigned by:

Hsin-ju Hsieh

814AF18B7A9147A...

21 May 2019

Hsin-ju Hsieh, Senior Director, Clinical Statistics

Statistical Analysis Plan

A Phase 1b/2 Study of MEDI4736 in Combination with Tremelimumab, MEDI4736 Monotherapy, and Tremelimumab Monotherapy in Subjects with Metastatic or Recurrent Gastric or Gastroesophageal Junction Adenocarcinoma

Protocol Number: D4190C00021

TABLE OF CONTENTS

███

███

███

LIST OF IN-TEXT TABLES

List of Abbreviations

Abbreviation or Specialized Term	Definition
ADA	antidrug antibody
AE	adverse event
AESI	adverse event of special interest
AUC	area under the concentration-time curve
BICR	blinded independent central review
CI	confidence interval
CR	complete response
CT	computed tomography
CTLA-4	cytotoxic T-lymphocyte-associated antigen 4
DC	disease control
DCR	disease control rate
DCR-16w	disease control rate at 16 weeks
DCR-24w	disease control rate at 24 weeks
DLT	dose-limiting toxicity
DoR	duration of response
DSD	duration of stable disease
EBV	Epstein-Barr virus
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
GEJ	gastroesophageal junction
irAE	immune-related adverse event
IRB	Institutional Review Board
irCR	immune-related complete response
irPD	immune-related progressive disease
irRECIST	immune-related Response Evaluation Criteria in Solid Tumors
IV	intravenous(ly)
IXRS	interactive response system
MTD	maximum tolerated dose
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NE	not evaluable
OR	objective response
ORR	objective response rate
OS	overall survival
PD	progressive disease
PD-L1	programmed cell death ligand 1
PFS	progression-free survival
PFS-6	PFS at 6 months

Abbreviation or Specialized Term	Definition
PFS-9	PFS at 9 months
PK	pharmacokinetic(s)
PR	partial response
Q2W	every 2 weeks
Q4W	every 4 weeks
QLQ-C30	Quality-of-Life Questionnaire Core 30
QLQ-STO22	EORTC symptom-specific module for gastric cancer
QTcB	QT corrected using Bazett's formula
QTcF	QT corrected using Fridericia's formula
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	stable disease
SID	Subject ID
SLD	sum of longest diameters
TTR	time to response

1 INTRODUCTION

This document describes the statistical analysis methodology for protocol D4190C00021, a randomized, multicenter, open-label, Phase 1b/2 study to evaluate the safety, tolerability, and clinical activity of MEDI4736 in combination with tremelimumab, MEDI4736 monotherapy, and tremelimumab monotherapy in subjects with metastatic or recurrent gastric or gastroesophageal junction (GEJ) adenocarcinoma. The main portion of this document details the statistical summaries relating to each study objective and describes 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(s)

1. <u>For Phase 1b</u>: To assess the safety and tolerability, describe any dose-limiting toxicities (DLTs), and determine the maximum tolerated dose (MTD) or the highest protocol-defined doses (in the absence of exceeding the MTD) for MEDI4736 in combination with tremelimumab in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma.

2. <u>For Phase 2</u>: To determine the objective response rate (ORR) and progression-free survival (PFS) at 6 months (PFS-6) based on Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 (Eisenhauer et al, 2009) as primary measures of clinical activity of MEDI4736 in combination with tremelimumab in second- and third-line subjects whose tumors have a positive interferon-gamma (IFN-γ) gene expression signature and in unselected second- and third-line subjects, MEDI4736 monotherapy, and tremelimumab monotherapy in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma.

2.1.2 Secondary Study Objectives

1. <u>For Phase 1b:</u> To assess disease control rate (DCR; defined as complete response [CR], partial response [PR], or stable disease [SD]), ORR, and PFS-6 based on RECIST v1.1 as measures of clinical activity of MEDI4736 in combination with tremelimumab in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma.

2. <u>For Phase 2</u>: To further describe the safety and tolerability of MEDI4736 in combination with tremelimumab, MEDI4736 monotherapy, and tremelimumab monotherapy in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma.

3. <u>For Phase 2</u>: To assess DCR and duration of response (DoR) based on RECIST v1.1, and overall survival [OS] as additional measures of clinical activity of MEDI4736 in combination with tremelimumab in second- and third-line subjects whose tumors have a positive IFN-γ gene expression signature and in unselected second- and third-line subjects, MEDI4736 monotherapy, and tremelimumab monotherapy in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma.

4. <u>For Phase 2</u>: To define the components of programmed cell death ligand 1 (PD-L1) expression (i.e., tumoral vs stromal) that correlate with clinical activity of MEDI4736 in combination with tremelimumab, MEDI4736 monotherapy, and tremelimumab monotherapy in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma.

2.2   Study Design

This is a randomized, multicenter, open-label, Phase 1b/2 study to evaluate the safety, tolerability, and clinical activity of MEDI4736 in combination with tremelimumab, MEDI4736 monotherapy, and tremelimumab monotherapy in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma who meet the protocol specified criteria.

The dose and schedule of MEDI4736 in combination with tremelimumab for Phase 1b of the current study will be based on emerging safety, clinical, and PK/pharmacodynamics data from the ongoing Study D4190C00006, which is evaluating the combination of MEDI4736 and tremelimumab in subjects with non-small-cell lung cancer (NSCLC). It is anticipated that a dose level and schedule determined as safe in NSCLC will also be safe in subjects with metastatic or recurrent gastric or GEJ adenocarcinoma. Two scenarios are listed in the protocol depending on the dose level and schedule decisions in study D4190C00006.

2.3   Treatment Assignment and Blinding

In Phase 1b, each subject who meets the eligibility criteria will be assigned open-label investigational product.

In Phase 2, subject who meets the eligibility criteria will be enrolled in 1 of 5 treatment arms. Second-line subjects with metastatic or recurrent gastric or GEJ adenocarcinoma will be randomized in a 2:2:1 ratio to Arm A (MEDI4736 in combination with tremelimumab), Arm B (MEDI4736 monotherapy), or Arm C (tremelimumab monotherapy). In parallel, third-line subjects with metastatic or recurrent gastric or GEJ adenocarcinoma will be enrolled in Arm D (MEDI4736 in combination with tremelimumab). As of Amendment 6, second-line and third-line subjects who have a positive IFN-γ gene expression signature as measured in archival formalin-fixed paraffin embedded (FFPE) tumor samples will be enrolled in Arm E (MEDI4736 20 mg/kg in combination with tremelimumab 1mg/kg Q4W). Of the first 20 subjects enrolled into Arm E, no more than 15 subjects may be third-line; no more than 30 subjects may be third-line in the overall cohort.

An interactive voice response system/web response system (IXRS) will be used to assign unblinded investigational product kit numbers. A subject is considered entered into the study when the investigator notifies the IXRS that the subject meets eligibility criteria and the IXRS provides the assignment of unblinded investigational product kit numbers to the subject. For subjects in the randomized portion of the study, the randomization code will be produced by an independent statistician, who is not part of the study team.

2.4   Sample Size

The planned sample size includes up to approximately up to 9 subjects in Phase 1b and up to approximately 126 subjects in Phase 2.

2.4.1 Phase 1b

A safety run-in of up to 9 subjects will be required to investigate the safety and tolerability of the dose level and schedule selected for dose expansion in Study D4190C00006. The sample size for Phase 1b of the study is not based on formal statistical power considerations.

2.4.2 Phase 2

<u>Arms A, B and C (Prior to Protocol Amendment 6)</u>

Approximately 125 subjects will be randomly assigned in a 2:2:1 ratio to Arms A, B, or C. Assuming 10% of the subjects are not evaluable for response or drop-out without a progression free survival (PFS) event before 6 months participation, approximately 45 evaluable subjects for either Arm A or B, and approximately 22 evaluable subjects for Arm C are expected. The sample size determination is based primarily on the consideration of providing the estimation with reasonable precision. As shown in Table 2.4.2-1, when the ORR is expected to be in the 29% to 40% range, a total of 45 subjects would provide a width of 11% to 12% between the observed point estimator and its lower limit of the exact 90% CI. If 24 of 45 subjects have PFS ≥ 6 months, the PFS-6 will have an estimate of 53% with a 90% CI of (41%, 65%). Assuming that PFS follows an exponential distribution, the PFS-6 of 53% will correspond to a median PFS of 6.6 months, which is a 50% increase from the 4.4 months observed for ramucirumab plus paclitaxel in subjects with metastatic gastric or GEJ adenocarcinoma following disease progression on first-line platinum- and fluoropyrimidine-based combination therapy in the RAINBOW study (Wilke et al, 2014)

Table 2.4.2-1   Observed ORR and PFS-6 with 90 Percent CI (Arms A and B)

N	Number of Subjects with OR	ORR	2-sided Exact 90% CI	
45	13	29.0%	18%	42%
45	18	40.0%	28%	53%

N	Number of Subjects with PFS at 6 Months	PFS-6	2-sided 90% CI	
45	24	53.0%	41%	65%

OR = objective response; ORR = objective response rate; PFS-6 = progression free survival 6 months

With 45 evaluable subjects for either Arms A or B in Phase 2, the study will have > 90% probability to select the superior regimen if the true difference between the response (or PFS-6) rates in those 2 treatment arms is ≥ 14%. In addition, assuming the prevalence of

PD-L1 expression is around 40%, approximately 18 evaluable subjects with high PD-L1 expression for either arms A or B will be expected. The probability of observing at least 1 responder out of 18 evaluable subjects will be at least 90% if the response rate is at least 15%. Additional subjects with high PD-L1 expression may be enrolled in order to further evaluate this population.

With approximately 45 evaluable subjects for Arm A and 22 evaluable subjects for Arm C, the study has 85% power to detect a difference of 35% vs 10% in the response (or PFS-6) rates assuming a 1-sided $\alpha = 0.10$ between the combination arm and the tremelimumab monotherapy arm.

Arm D

It is planned to enroll 25 subjects in Arm D.

Arm E (as of Protocol Amendment 6)

Approximately 40 additional subjects who are biomarker positive will be assigned to Arm E. Assuming 10% of the subjects are not evaluable for response or drop-out without PFS event before 6 months, approximately 36 evaluable subjects for Arm E are expected. The sample size determination is based primarily on the consideration of providing the estimation with reasonable precision. As shown in Table 2.4.2-2, when the ORR is expected to be in the 28% to 39% range, a total of 36 subjects would provide a width of 12% to 14% between the observed point estimator and its lower limit of the exact 90% CI. If 19 of 36 subjects have PFS $\geq$ 6 months, the PFS-6 will have an estimate of 53% with a 90% CI of (39%, 67%). Assuming that PFS follows an exponential distribution, the PFS-6 of 53% will correspond to a median PFS of 6.6 months, which is a 50% increase from the 4.4 months observed for ramucirumab plus paclitaxel in subjects with metastatic gastric or GEJ adenocarcinoma following disease progression on first-line platinum- and fluoropyrimidine-based combination therapy in the RAINBOW study.

Table 2.4.2-2   Observed ORR/PFS-6 with 90% CI (Arm E)

N	Number of Subjects with OR	ORR	2-sided Exact 90% CI	
36	10	28%	16%	43%
36	14	39%	25%	54%

N	Number of Subjects with PFS at 6 Months	PFS6	2-sided 90% CI	
36	19	53%	39%	67%

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). 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 summarized by treatment group. Treatment group refers to dose cohorts in phase 1b and treatment arms in phase 2.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.

Two-sided confidence intervals, whenever specified, will be produced at 90% and 95% significance levels.

In general, the baseline is defined as the last non-missing observation collected on or prior to the date of the first dose of study drug(s) for treated subjects unless stated otherwise.

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 SAS® programs used to generate analytical results will be developed and validated according to MedImmune SAS® programming standards and MedImmune SAS® validation procedures.

3.2 Analysis Populations

The analysis populations are defined in Table .

Table 3.2-1   Analysis Populations

Population	Description
As-treated population	Subjects who receive any study investigational product will be included in the as-treated population and subjects will be analyzed according to the treatment they actually receive.
Response evaluable population	The response evaluable population includes all subjects from the As-treated Population who have a baseline disease assessment with measurable disease and one of the following: (1) at least one post-baseline disease assessment, and/or (2) withdrawn from study treatment prior to post-baseline disease assessment due to clinical progression or death.
DLT evaluable population	The DLT evaluable population includes all subjects enrolled in Phase 1b who receive the protocol-assigned treatment and complete safety follow-up through the DLT evaluation period or experience a DLT during the DLT evaluation period.

Table 3.2-1   Analysis Populations

Population	Description
Crossover population	Crossover population includes all subjects who crossover to receive combination therapy (MEDI4736 in combination with tremelimumab) after confirmed progression on the initial treatment assignment of MEDI4736 or tremelimumab monotherapy.
Re-treated population	Re-treated population includes all subjects who have been re-treated with combination therapy (MEDI4736 in combination with tremelimumab) during the first 12 months of follow-up.

If the number of subjects in the re-treatment population or crossover population is too small (e.g. ≤5), the planned analysis for the re-treatment phase may be reduced and simplified. Details will be presented in the statistical programming plan (SPP).

3.3   Study Subjects

3.3.1   Subject Disposition and Completion Status

A summary of subjects treated at each dose level in Phase 1b as well as subjects randomized or treated at each treatment arm in Phase 2 will be provided. In addition, summaries using the number and percent of subjects that completed or discontinued treatment and the reasons for discontinuation during the initial treatment and re-treatment will be provided.

Summaries using the number and percent of subject status at the end of study and the reasons for ending study will be provided. 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 (e.g., toxicity related to investigational product or disease under investigation).

3.3.2   Demographics and Baseline Characteristics

Demographic information and baseline disease characteristics will be summarized for the As-treated population. Demographic information related to sex, age, race, ethnicity, weight, height, and body mass index will be summarized. Tumor diagnosis including disease stage, extent of disease at study entry, time from initial diagnosis to study entry, MSI testing gene status (present/absent), EBV evaluation of the tumor status (positive/negative), helicobacter pylori status (positive/negative), anatomic location of primary tumor, histology subtype and number of organs involved with metastasis will be summarized.

The summary for prior anticancer treatment will include the number and percent of subjects by: lines of therapy for recurrent/metastatic disease, treatment setting, type of treatment (i.e. biologic, immunotherapy, radiotherapy, chemotherapy, surgery, and other), and best response (complete response, partial response, stable disease, progressive disease, not evaluable, and not applicable) to the most recent line of therapy.

Baseline tumor characteristics including number and sites of target and non-target tumor lesions as well as tumor burden (sum of target lesion diameters) will be summarized.

3.3.3 Study Drug Exposure

The number of MEDI4736 and tremelimumab doses/cycles and total MEDI4736 and tremelimumab duration of exposure will be summarized by descriptive statistics and by frequency. ████████████████████████

██

████████

██

██

████████

Number of doses is defined as the total number of doses received. A dose will be counted if treatment is started even if the full dose is not delivered. The dose intensity (mg/kg/cycle length) is defined as total actual dose (in mg/kg) that a subject received per study defined cycle. Relative dose intensity is defined as the dose intensity divided by the planned dose per cycle. The details of the dose intensity calculation will be provided in the SPP.

Dosing delays, interruptions and omissions of MEDI4736 and tremelimumab will be derived based on the scheduled dosing dates and will be summarized.

3.3.4 Concomitant Medications

The number and percentage of subjects who took at least 1 dose of medication other than investigational product(s) during the study will be summarized by the ATC class and preferred term coded by AstraZeneca Drug Dictionary (AZDD) 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 investigational products or any concomitant medication started prior to first dose of study treatment that continued beyond the date of first dose of investigational products.

3.4   Efficacy Analyses

The efficacy analysis will be based on the As-treated Population. Sensitivity analyses for response-related endpoints (OR, DoR, and DCR) will be performed based on the Response Evaluable Population.

Primary analyses for response-related endpoints and corresponding time-to-event endpoints for primary and secondary efficacy endpoints will be based on the programmatically-derived response from the investigator's recorded measurements and assessments for target, non-target, and new lesions according to RECIST v1.1 or blinded independent central review (BICR) if performed.

3.4.1   Primary Efficacy Endpoints and Analyses

Arms A, B, C and E

The two co-primary efficacy endpoints are Objective response (OR) and PFS-6 based on RECIST v1.1.

Arm D

The primary efficacy endpoint is objective response (OR) based on RECIST v1.1.

The primary efficacy analyses for the primary efficacy endpoint(s) will be based on the As-treated Population.

Objective Response

Objective response (OR) is defined as best overall response (BOR) of confirmed CR or confirmed PR according to RECIST v1.1. The best overall response is defined as the best response (in the order of CR, PR, SD, PD, and not evaluable) among all overall responses recorded from the date of randomization for Phase 2 Arm A, B and C subjects or the date of first dose of study treatment for Phase 1b and Phase 2 Arm D and E subjects until progression, or the last evaluable disease assessment in the absence of PD prior to the initiation of subsequent anticancer therapy or discontinuation from the study, whichever occurs first.

The best overall response (BOR) of CR or PR must be confirmed, which means a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging not less than 28 days (4 weeks) after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visit. A confirmed CR is defined as two CRs that were

separated by at least 28 days with no evidence of progression in-between. A confirmed PR is defined as two PRs or an un-confirmed PR and an un-confirmed CR that were separated by at least 28 days with no evidence of progression in-between. For a subject to qualify for a BOR of SD, at least 54 days from the randomization date or start of treatment date must elapse without disease progression. The definition of 54 days reflects the protocol-defined disease assessment window of ±3 days (Day 57 plus/minus 3 days).

Subjects who responded with an unconfirmed CR/PR at the time of data cutoff will be reported as CR/PR pending confirmation under the SD category provided the minimum criteria for SD duration (54 days) are met, otherwise this will be reported under the Non Evaluable (NE) category. NE is assigned for a subject who has only a baseline assessment, or a response assessment of CR/PR/SD at an interval of less than 54 days since randomization or the start of treatment and who has no subsequent disease evaluation, or has all overall response evaluations assessed as NE.

The ORR is defined as the proportion of subjects with OR. The 2-sided 90% and 95% CI of ORR will be estimated based on the exact probability method. In addition, to help make a comparison to historical data, the unconfirmed ORR which is defined as the proportion of subjects who achieved a best overall response of confirmed/unconfirmed CR or confirmed/unconfirmed PR will be presented along with its CI.

ORR will be compared between the three treatment groups (Arm A, B and C). P-values will be presented without multiplicity adjustment. The comparisons will include the following:

- MEDI4736 + tremelimumab combination therapy (Arm A) versus MEDI4736 monotherapy (Arm B)

- MEDI4736 + tremelimumab combination therapy (Arm A) versus tremelimumab monotherapy (Arm C)

P-values for comparisons of ORR between the treatment groups mentioned above will be obtained from Fisher's exact test. The statistical significant treatment difference will be tested against 2-sided alpha level of 0.05. In addition, the difference in ORR between MEDI4736 + tremelimumab combination and MEDI4736 monotherapy or tremelimumab and associated 95% exact unconditional confidence interval will be provided using Agresti and Ming (2001)'s approach by specifying the RISKDIFF(METHOD=FMSCORE) option in PROC FREQ.

PFS-6

PFS-6 is 6-month progression-free survival rates, which is the proportion of subjects who are progression free and alive at 6 months. PFS is defined as the time from the date of randomization for Phase 2 Arm A, B and C subjects or the date of first dose of study treatment for Phase 1b and Phase 2 Arm D and E subjects to the earlier of the dates of the first objective documentation of radiographic disease progression (per RECIST v1.1) or death due to any cause.

For subjects who are alive with no objective documentation of (radiographic) disease progression by the data cutoff date, PFS will be censored at the date of their last evaluable tumor assessment. If the subject progresses or dies after ≥ 2 consecutively missed or non-evaluable disease assessment visits, PFS will be censored at the time of the latest evaluable disease assessment prior to the missed assessments.

In this study, progression or death after ≥2 consecutive missed disease assessment visits is defined (according to disease assessment schedules specified in the protocol) as below:

- ≥ 18 weeks (two disease assessment visits plus 2 week visit window to allow for a late assessment) since the last post-baseline evaluable disease assessment, if progression or death occurs prior to end of treatment.
- ≥ 26 weeks (three disease assessment visits plus 2 week visit window to allow for a late assessment) since the last post-baseline evaluable disease assessment, if progression or death occurs after end of treatment.

The details of censoring rules are listed in table 3.4.1-1.

Table 3.4.1-1   Summary of Censoring Guidelines for PFS

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 <=1 missed or non-evaluable disease assessment following an adequate post-baseline disease assessments	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/the first dose of study treatment or last progression-free disease assessment prior to missed or non-evaluable assessments, whichever occurred last	Censored
Initiation of alternative anticancer therapy before PD or death	Date of last adequate progression-free disease assessment prior to initiation of alternative anticancer therapy	Censored
No tumor assessment at baseline and no death prior to second scheduled post-baseline disease assessment, or no tumor assessment post-first dose	Date of randomization/the first dose of study treatment	Censored

Kaplan-Meier curves will be generated for PFS and used to estimate the PFS-6 along with its 2-sided 90% and 95% CIs. The CIs for PFS-6 will be calculated by applying asymptotic normality to the log-log transformation of PFS-6.

PFS-6 will be compared between the three treatment groups (Arm A, B and C). P-values will be presented without multiplicity adjustment. The comparisons will include the following:

- MEDI4736 + tremelimumab combination therapy (Arm A) versus MEDI4736 monotherapy (Arm B)

- MEDI4736 + tremelimumab combination therapy (Arm A) versus tremelimumab monotherapy (Arm C)

The comparison of PFS-6 will be performed by normal approximation under log-log transformation of Kaplan-Meier estimates of PFS-6 (Klein et al, 2007).

Specifically, the log HR (1 for combination vs 2 for monotherapy) is estimated as

$$\log(-\log(\hat{S_1}(t))) - \log(-\log(\hat{S_2}(t)))$$

The variance for log HR is estimated as

$$\frac{Var(\hat{S}(t))}{\hat{S_1}(t)(\log(\hat{S}(t)))^2} + \frac{Var(\hat{S}(t))}{\hat{S_2}(t)(\log(\hat{S}(t)))^2}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimate of the progression free survival function, $Var_i(\hat{S_i}(t)) = \hat{S_i}(t)^2 \sum_i \frac{d_{ik}}{t_{ik} \leq t \, n_{ik}(n_{ik}-d_{ik})}$ is the variance for $\hat{S}_i(t)$ derived from Greenwood's formula. d_{ik} and n_{ik} refer to the number of events and number of subjects at risk in the ith group at time t_{ik}, respectively. The final estimate of HR and corresponding 95% CI will be obtained by taking the exponential of log HR and its 95% confidence limits.

3.4.2   Secondary Efficacy Endpoints and Analyses

Phase 1b

Secondary efficacy endpoints include OR and disease control (DC), duration of stable disease (DSD), best percentage change from baseline of the sum of the longest diameters (SLD) of target lesions and PFS-6.

Phase 2

Secondary efficacy endpoints include duration of response (DoR), time to response (TTR), DC, duration of stable disease (DSD), best percentage change from baseline of the sum of longest diameters (SLD) of target lesions, PFS, and overall survival (OS).

The definitions of these efficacy endpoints and associated analyses are described below. The primary efficacy analyses for the secondary efficacy endpoints will be based on the As-treated Population.

* Disease Control (DC) is defined as a best overall response of confirmed CR, confirmed PR or SD per RECIST v1.1. Disease assessments at Week 16 and 24 for this study are performed on Day 113 (+/- 3 days) and Day 169 (+/- 3 days), therefore the disease control rate at 16 and 24 weeks (DCR-16w and DCR-24w) for this study is defined as the proportion of subjects who achieved a best overall response of confirmed CR, confirmed PR or have SD with duration of SD for a minimum duration of 110 days and 166 days, following the date of randomization for Phase 2 Arm A, B and C subjects or the date of first dose of study treatment for Phase 1b and Phase 2 Arm D and Arm E subjects, respectively. The 2-sided 90% and 95% CIs for DCR-16w and DCR-24w will be estimated using the exact binomial distribution.

- The DoR is defined as the time from the date of first documented response (CR or PR) until the first date of documented progression according to RECIST v1.1 that occurs subsequently after response or death due to any cause, whichever occurs first. For subjects who are alive and progression-free at the time of data cut-off for analysis, DoR will be censored at the PFS censoring time. The DoR will only be evaluated for subjects with an OR and will be calculated using the Kaplan-Meier method.

- TTR is defined as the time from the date of randomization for Phase 2 Arm A, B, and C subjects or the date of first dose of study treatment for Phase1b and Phase 2 Arm D and Arm E subjects until the first documented OR per RECIST v1.1 and will be evaluated only for subjects with an OR. For subjects proceeding from PR to CR, the onset of PR is taken as the onset of response. TTR will be summarized by treatment group with descriptive statistics.

- DSD is defined as the time from the date of randomization for Phase 2 Arm A, B, and C subjects or the date of first dose of study treatment for Phase 1b and Phase 2 Arm D and Arm E subjects until the first date of documented PD (per RECIST v1.1), or death due to any cause, whichever occurs first. DSD will be analyzed for subjects with SD as their BOR per RECIST v1.1. The median DSD with 90% and 95% CI will be estimated based on the Kaplan-Meier curves. For subjects who are alive and progression-free at the time of data cut-off, DSD will be censored at the PFS censoring time.

- Best percentage change from baseline of the SLD of target lesions per RECIST v1.1 will be derived as the biggest decrease or the smallest increase from baseline on the SLD among all post-baseline disease assessment including unscheduled assessments. Best percentage change from baseline in the SLD of target lesions by subject will be presented in a 'Waterfall' plot, to show each subject's best percentage as a separate bar, with the bars ordered from the largest increase to the largest decrease. Additionally, the percent change from baseline in the SLD of target lesion at each post-baseline disease assessment will be summarized by treatment group with descriptive statistics.

- PFS is defined as in section 3.4.1.2. PFS-9 is the 9-month PFS rate, which is the proportion of subjects who are progression free and alive at 9 months. Similar to PFS-6, Kaplan-Meier curves will be generated for PFS and used to estimate median PFS and the PFS-9 rate along with their 2-sided 90% and 95% CIs.

- Overall survival is defined as the time from date of randomization for Phase 2 Arm A, B, and C subjects or the date of first dose of study treatment for Phase 2 Arm D and Arm E

subjects until death due to any cause. If there is no death reported for a subject by the data cut-off date, OS will be censored at last known alive date. If the last known alive date is after the data cutoff date, OS will be censored at the date of data cutoff for that subject. Kaplan-Meier curves will be generated for OS and used to estimate the median OS and the 1-year survival rate along with their 2-sided 90% and 95% CIs.

3.4.3   Subgroup Analyses

In order to assess the consistency of treatment effect, the efficacy analyses of ORR, DCR-16w, DCR-24w, DoR, PFS, and PFS-6 will be summarized based on the As-treated population for the following subgroups provided sufficient data are available:

- PD-L1 status (positive, negative, unselected)

- Age (<65, $\geq$ 65 yrs.)

- Region (US, rest of world)

- Race (Asian, non-Asian)

3.4.4   Other Efficacy Analyses

In addition, sensitivity analyses for response-related endpoints and corresponding time-to-event endpoints for primary and secondary efficacy endpoints will be based on overall response reported based on investigator assessment if there is sufficient discrepancy.

Sensitivity analyses for ORR, DCR-16w and DCR-24w will be performed based on the Response Evaluable Population.

Sensitivity analyses for the ORR, DCR-16w, DCR-24w, DoR, PFS, PFS-6, and PFS-9, all derived according to RECIST v1.1 modified for confirmation of PD may be performed. Under RECIST v1.1 with modification, any objective disease progression must be confirmed by the next scheduled scan. The confirmatory scan must be no sooner than 4 weeks after the initial suspected progression. If disease progression is confirmed (or disease progression occurs and no further scans are recorded) then the date of progression will be when it was originally observed. Subjects with a single disease progression and no further tumor assessment scans will be treated as PD in the analysis. For such sensitive analysis, tumor data obtained up until confirmed progression, or the last evaluable assessment in the absence of a confirmed progression (prior to starting new anticancer therapy), will be included in the

assessment of BOR, ORR and DCR. Note that the response or disease control may occur after an unconfirmed progression.

[REDACTED]

3.4.5 Handling of Dropouts and Missing Data

Missing target lesion measurement data

For the disease response derivation in target lesion (TL) data, if a lesion size is missing due to the lesion being too small to be measured, the missing value is defaulted to 5 mm. If a lesion size is missing due to technical difficulty of the measurement (i.e., poor quality of scan, blurring lesion etc.), the measurement for that lesion will not be imputed and the target lesion visit response is not evaluable (NE). Overall visit response will also be NE. However, a visit response of PD should still be assigned if any of the following occurred:

- A new lesion is recorded
- A non-target lesion (NTL) visit response of PD is recorded
- The sum of TLs is sufficiently increased to result in a 20% or greater increase, and an absolute increase of ≥ 5mm, both from nadir even assuming the non-recorded TLs have disappeared

3.5 Patient Reported Outcomes

3.5.1 EORTC QLQ-C30 and Analysis of EORTC QLQ-C30

The EORTC QLQ-C30 v3.0 is a 30-item self-administered questionnaire. EORTC QLQ-C30 is measured at baseline, during treatment and during the follow-up period. The scoring of the symptom scales/items will follow the EORTC scoring manual. Individual items will be assessed using untransformed scores but for each of the 15 domains (9 multiple-item scales, 6 single-item scales), final scores are transformed such that they range from 0-100; higher scores indicate greater functioning, greater quality of life, or greater level of symptom burden (Aaronson et al, 1993).

For each subscale, if less than 50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales. If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing. The reason for any missing questionnaire will be identified and recorded. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimized.

Table 3.5.1-1   Time to deterioration of Health Related QoL

Score	Change from baseline	Visit response
EORTC QLQ-C30 multiple-item scales: Global HRQoL score; functional scale (physical, role, emotional, cognitive, and social) scores; symptom scale (fatigue, pain, nausea & vomiting) scores	Increase of 10 points or more Decrease of 10 points or more Otherwise	Improved Deterioration No change
EORTC QLQ-STO22 Symptom scores: STODYS (dysphagia) score, STOEAT (eating restrictions) score, STOPAIN (stomach pain) score, STOFX (reflux) score,	Decrease of 10 points or more Increase of 10 points or more Otherwise	Improved Deterioration No change

The deterioration must be sustained for at least 21 days (i.e. there must be no response of "improved" or "no change" within 21 days of the visit response of "deterioration").
Higher scores on the functional scales indicate a higher level of functioning and on the global HRQoL score indicate a better quality of life. Higher scores on the symptom scales/items indicate a higher symptom burden.

Time to deterioration of will be calculated for the multiple-item scales: 5 functional scales (physical, role, emotional, cognitive and social); 3 multi-item symptom scales (fatigue, pain, nausea and vomiting) and a 2-item global HRQoL scale. A change of at least 10 points in the score will be considered as a clinically relevant or a minimally important difference (Table 3.5.1-1).

Time to deterioration of a multiple-item scale will be defined as the time from date of randomization for Phase 2 Arm A, B, and C subjects or the date of first dose of study treatment for Phase 1b and Phase 2 Arm D and Arm E subjects to the date of a clinically important deterioration or death (by any cause) in the absence of a clinically meaningful deterioration regardless of whether the patient withdraws from treatment or receives another anti-cancer therapy prior to deterioration in the multiple-item scale score. Death will be included as an event only if the death occurs within two visits of the last PRO assessment

where the symptom change could be evaluated.

Subjects whose multiple-item scale score has not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last EORTC assessment where the symptom could be evaluated. Also, if the multiple-item scale score deteriorates after 2 or more missed EORTC assessment visits or the patient dies after 2 or more missed EORTC assessment visits, the patient will be censored at the time of the last EORTC assessment where the multiple-item scale score could be evaluated. Time to deterioration of the multiple-item scales will be analyzed using the same methodology and model as described for the analysis of PFS.

Summary tables of the absolute change from baseline for domains and subscale scores will be summarized descriptively.

Compliance overall and over time will be summarized by treatment groups for the EORTC QLQ-C30 questionnaire. These will be based upon:
- Received forms = number of EORTC QLQ-C30 forms received back.
- Expected forms = number of subjects still under QoL follow-up at the specified assessment time excluding subjects in countries with no available translation. For subjects whose disease has progressed, the latest of progression and safety follow-up will be used to assess whether the patient was under QoL follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms. Evaluable forms = subset of expected EORTC QLQ-C30 forms with at least one subscale that can be determined

Overall compliance will be defined as the number of subjects who provided both a baseline and at least one post baseline assessment for which there were sufficient data recorded for the visit to be evaluable for at least one multi-item or single-item scale score, divided by the number of subjects treated/randomized. Compliance over time is calculated separately for each visit, including at baseline, as the number of subjects with an evaluable form at the time point (as defined above), divided by number of subjects expected to have completed forms at that visit. Similarly the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable forms (per definition above), divided by the number of received forms.

Best QoL response per subject overall is defined as shown in Table 3.5.1-2 below.

Table 3.5.1-2   Best QoL response

Overall score response	Criteria
Improved	Two visit responses of "improved" a minimum of 21 days apart without an intervening visit response of "deterioration"
No change	Does not qualify for overall score response of "improved". Two visit responses of either "no change" or "improved" and "no change" a minimum of 21 days apart without an intervening visit response of "deterioration"
Deterioration	Does not quality for overall score response of "improved". A visit response of "deterioration" without a response of "improved" or "no change" within 21 days
Other	Does not quality for one of the above

3.5.2   EORTC QLQ-STO22 and Analysis of EORTC QLQ-STO22

The EORTC QLQ-STO22, a disease-specific 22-item self-administered questionnaire for gastric cancer (Vickery et al, 2001), was developed to be used in conjunction with the EORTC QLQ-C30 v3.0. The scoring of the symptom scales/items will follow the EORTC scoring algorithm. For the symptom scales and items a higher score is equivalent to worse or more symptoms. In the functional scales, a high score is equivalent to better function.

Time to deterioration analysis will be conducted in the same for the following symptom scales of the EORTC QLQ-STO22 as described for the EORTC QLQ-C30 multiple-item scales: STODYS (dysphagia) score, STOEAT (eating restriction) score, STOPAIN (stomach pain) score, and STOFX (reflux) score. A change of at least 10 points in the score will be considered as a clinically relevant or a minimally important difference.

Change from baseline for each of the summary scales and single item measures described above will be summarized descriptively. Summary symptom scores: STODYS (dysphagia) score, STOEAT (eating restriction) score, STOPAIN (stomach pain) score, and STOFX (reflux) score will be included. Multiple item system scale STOANX (anxiety) and single item measures: STODM (dry mouth) score, STOT (taste) score, STOBI (body image) score, and STOHL (hair loss) score will not be summarized but will be presented in by-subject listings.

Similarly, compliance overall and over time will be summarized by randomized or assigned treatment for the EORTC QLQ-STO-22 questionnaire as was described for the EORTC QLQ-C30.

3.6 Pharmacodynamic Endpoint(s) and Analyses

3.6.1 Pharmacodynamic Endpoint(s)

Exploratory pharmacodynamic endpoints are listed in Section 2.2.3 of the protocol.

3.6.2 Analysis of Pharmacodynamic Endpoint(s)

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

3.7 Other Additional Analyses

Not applicable.

3.8 Safety Analyses

In Phase 1b, the number of DLTs that are identified in the DLT Evaluable Population during the DLT evaluation period will be summarized and listed. Safety data including AEs, SAEs, laboratory parameters, ECGs, vital signs and ECOG performance status, will be summarized in the As-treated population by treatment group, pooled dose groups and overall.

Repeated or unscheduled tests will not be summarized for each scheduled visit, but will be included for summaries of "worst-case" on treatment values and shift-table analyses. "Worst-case" on treatment (i.e. maximum or minimum on-treatment value depending on the direction of an adverse effect) is defined as the nadir and/or zenith including any scheduled and unscheduled post-baseline assessments that occur through the last assessment on-treatment. The last assessment on-treatment is defined as the last visit with a non-missing observation that occurred up till 90 days (+7 days to allow for a visit window) following the last dose of study treatment (if safety follow-ups are performed till 90 days after the last dose), or till the initiation of the subsequent anticancer therapy (excluding palliative radiotherapy), whichever occur first.

If the number of subjects with available data is less than 10% of the total number of subjects in the As-treated Population for overall summary for the dose escalation phase, or for each cohort in the dose expansion phase at a scheduled time of evaluation, no summary statistics will be presented at that time point, unless otherwise indicated.

3.8.1 Adverse Events and Serious Adverse Events

Adverse events will be coded using Medical Dictionary for Regulatory Activities and assigned grades based on NCI CTCAE v4.03.

Treatment-emergent adverse event (TEAE) will be defined for the initial treatment period and subsequent treatment periods separately. TEAEs are defined as any adverse events that occur on or after the date of initial receipt of study treatment during each treatment period. Analysis of AEs (as described below) will be limited to TEAEs. Any AEs that are considered as non-treatment emergent will be presented in the listings only.

For both Phase 1b and Phase 2, the number and percentage of subjects reporting treatment-emergent AEs will be summarized overall and by the worst NCI CTCAE v4.03 grade, system organ class, and preferred term. Similarly, the number and percentage of subjects reporting treatment-emergent AEs considered related to investigational product will be summarized. At each level of subject summarization, a subject will be counted once using the highest grade and level of causality if one or more occurrences of the same system organ class/preferred term is reported. SAE will be summarized overall, by system organ class and preferred term, and by SAE criteria. The number and percentage of subjects reporting treatment-emergent SAEs considered related to investigational product will also be summarized.

3.8.2   Adverse Events of Special Interest

Some clinical concepts (including some selected individual preferred terms) have been considered "AEs of special interest" (AESI) to the MEDI4736 program. Adverse events of special interest are defined in the protocol section 5.3. The number and frequency of subjects reporting AESI will be summarized.

3.8.3   Deaths and Treatment Discontinuations due to Adverse Events

The number and percentage of subjects reporting treatment-emergent AEs leading to treatment discontinuation and death will also be summarized in a similar format as SAE.

3.8.4   Clinical Laboratory Evaluation

Laboratory tests will be grouped according to hematology, serum chemistry, urinalysis, coagulation, and thyroid function tests (TSH, free T3, and free T4). For all continuous laboratory assessments, absolute value and change from baseline will be summarized for each scheduled assessment, "worst-case" (nadir and/or zenith) on treatment, and the last assessment on-treatment using descriptive statistics.

Laboratory abnormalities with toxicity grades according to the NCI CTCAE 4.03 will be derived according to laboratory values. Frequencies of worst observed Grades, as defined by NCI CTCAE v4.03, will be presented for each laboratory parameter. The analysis will present the rates of subjects with Grade 3-4 toxicity. For each lab parameter, percentages are

calculated based on the number of subjects who have a baseline and at least one post-baseline assessment.

Also, laboratory parameters will be assessed by presenting tables containing information related to 2-grade (or greater) laboratory shifts from baseline.

3.8.5   Other Safety Evaluations

3.8.5.1   Vital Signs

Vital signs will be measured on study days noted in Protocol Section 4.2. Descriptive statistics of baseline value, post-baseline value and change from baseline value for heart rate, blood pressure, temperature, and respiratory rate will be provided for each scheduled assessment including the last assessment on-treatment as well as for the maximum and minimum post-baseline on treatment values.

3.8.5.2   Electrocardiogram

Electrocardiogram parameters (PR, RR, QRS, QT, QT corrected using Fridericia's formula QTcF) will be summarized using descriptive statistics for actual values and for changes from baseline by treatment arm by scheduled time of evaluation including end of treatment visit as well as for the maximum post-baseline values. The QTcF will be considered as the primary correction method to assess subject cardiac safety.

The notable ECG interval values in maximum absolute QTcF intervals (new > 450 milliseconds, new > 480 milliseconds, new > 500 milliseconds) and the maximum absolute uncorrected QT intervals (new > 500 milliseconds) over all post-baseline evaluations, as well as in QTcF maximum changes from baseline (> 30 and > 60 milliseconds) over all post-baseline evaluations will be summarized by treatment. "New" means the category of the QTc abnormality was not present at baseline and became present at least one post-baseline ECG assessment.

3.8.5.3   ECOG Performance Status

ECOG performance status will be summarized using a shift table showing change in ECOG from baseline to the worst performance status on-treatment and to the last assessment on-treatment.

3.8.6   Subgroup Analyses

Not applicable.

5 REFERENCES

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; 19:3936-43.

Wilke et al. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in subjects with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. Lancet Oncology. 2014

A. Agresti and Y. Min. On small-sample confidence intervals for parameters in discrete distributions, Biometrics (2001) (A. Agresti and Y. Min) Biometrics. 2001 Sep;57(3):963-71.

Klein JP, Logan B, Harhoff M, Anderson PK. Analyzing survival curves at a fixed point in time. Statistics in Medicine. 2007; 26:4505-19.

6 APPENDIX

Table A. Statistical Analysis Plan Amendments

SAP Version	Date	Key Details of Amendment	Reason for Amendment
1.0	23MAR2016	Initial document	Initial document
2.0	18MAY2017	• Added the sample size calculation for the new Arm E (Section 2.4.2) • Analyses related to the new Arm E and some clarifications added in the efficacy analysis section (Section 3.4)	• To be aligned with protocol language
2.1	17APRIL2019	• Correct exact and standard confidence intervals for ORR and PFS (Section 2.4.2) • Inclusion of Arm E in the definition of OR. Correction of duration of non-treatment interval following 3 missed doses (Section 3.4.1) • Inclusion of Arm E in definition of secondary efficacy endpoints. Correct method of assignment of Arm E cohort (Section 3.4.2) • Correct method of assignment of Arm E cohort. Include missing multiple-item axis and single item STO22 scores for EORTC (Section 3.5.1) • Editorial changes for clarity (multiple sections)	• Insure methods completeness for CSR development • To be aligned with protocol language • To be aligned with protocol language • To be aligned with protocol language and EORTC scoring methodology • To insure proper presentation of final CSR TFLs
3.0	15MAY2019	• Remove multiple-item STOANX endpoint and STO22 single item endpoints from QoL summaries	• No clinical meaningfulness for these endpoints.