AstraZeneca

Statistical Analysis Plan

Study Code	D4198C00001
Edition Number	2.0
Date	27/Feb/2017

A Phase II Open-Label, Multi-Center Study of MEDI4736 Evaluated as Single Agent or in Combination with Tremelimumab in Patients with Metastatic Pancreatic Ductal Adenocarcinoma

A Phase II Open-Label, Multi-Center Study of MEDI4736 Evaluated as Single Agent or in Combination with Tremelimumab in Patients with Metastatic Pancreatic Ductal Adenocarcinoma

PPD

PPD

Study Statistician

3-Mar-2017
Date

A Phase II Open-Label, Multi-Center Study of MEDI4736 Evaluated as Single Agent or in Combination with Tremelimumab in Patients with Metastatic Pancreatic Ductal Adenocarcinoma

Global Product Statistician PPD

3-MAR-2017
Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

The following abbreviations and special terms are used in this study analysis plan.

Abbreviation or special term	Explanation
5-FU	Fluoropyrimidine
ADA	Antidrug antibody
AE	Adverse Event
AESI	Adverse Events of Special Interest
ALT	Alanine aminotransferase
ALP	Alkaline phosphatase
AST	Aspartate aminotransferase
BICR	Blinded Independent Central Review
BoR	Best objective Response
CI	Confidence Interval
CR	Complete Response
CSP	Clinical Study Protocol
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Event
DBL	Database lock
DCO	Data cut-off
DCR	Disease Control Rate
DCR12	Disease Control Rate at 12 weeks
DNA	Deoxyribonucleic acid
DoR	Duration of Response
d.p	Decimal place
DRM	Data Review Meeting
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EORTC	European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer quality of life core questionnaire

Abbreviation or special term	Explanation
EORTC QLQ-PAN26	European Organisation for Research and Treatment of Cancer quality of life questionnaire pancreatic cancer module
EQ-5D-5L	EuroQol 5-dimension, 5-level health state utility index
FAS	Full Analysis Set
fT_3	Free triiodothyronine
fT_4	Free thyroxine
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
IHC	Immunohistochemical
IP	Investigational Product
irAE	Immune-related adverse event
irRECIST	Immune-related Response Evaluation Criteria in Solid Tumors
ITT	Intent-to-treat
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
MDSC	Myeloid-derived suppressor cells
MedDRA	Medical Dictionary for Regulatory Activities
miRNA	Micro-ribonucleic acid
MRI	Magnetic Resonance Imaging
mRNA	Messenger ribonucleic acid
NE	Not evaluable
NTL	Non-target lesion
OAE	Other Significant Adverse Event
ORR	Objective Response Rate
OS	Overall Survival
OS6	Proportion of patients alive at 6 months from randomization/enrollment
OS12	Proportion of patients alive at 12 months from randomization/enrollment
PD	Progressive Disease
PDx	Pharmacodynamics
PDAC	Pancreatic ductal adenocarcinoma
PD-L1	Programmed cell death ligand 1
PFS	Progression-free survival

Abbreviation or special term	Explanation
PFS3	Proportion of patients with progression-free survival after 3 months
PFS6	Proportion of patients with progression-free survival after 6 months
PGx	Pharmacogenetic
PI	Principal Investigator
PID	Percentage intended dose
PK	Pharmacokinetics
PR	Partial Response
PRO	Patient Reported Outcomes
PRO-CTCAE	Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events
q4w	Every 4 weeks
q6w	Every 6 weeks
q12w	Every 12 weeks
QoL	Quality of Life
QLQ	Quality of Life Questionnaire
QTcF	QT interval corrected for heart rate using Fridericia's formula
RCT	Randomized Controlled Trial
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
SNP	Single Nucleotide Polymorphism
SoC	Standard of Care
sPD-L1	Soluble programmed cell death ligand 1
TFST	Time to First Subsequent Therapy
TSH	Thyroid Stimulating Hormone
ULN	Upper Limit of Normal

AMENDMENT HISTORY

Date	Brief description of change
V1.0 (16FEB2016)	N/A
V2.0 (27/Feb/2017)	Part B related revision and other minor revisions

1. STUDY DETAILS

This study consists of Part A, Lead-in, as well as a possible Part B, expansion (Part B: Expansion) and/or a possible Part B, randomized controlled study (Part B: RCT).

The purpose of this Statistical Analysis Plan (SAP) is to detail statistical methods and presentations for Part A: Lead-in and Part B: expansion only.

The format of Part B will be determined based on the responses seen in the first 30 patients in each arm in Part A. In the event a Part B: Expansion is initiated, data from Part A and Part B: Expansion will be combined. In the event a Part B: RCT is initiated, a separate SAP will be produced to detail the methods that will be applied.

This document has been prepared in conjunction with the Clinical Study Protocol (CSP) version 1 (17AUG2015) and is referring to the unique CRF template version 2.0 (05NOV2015).

A protocol amendment will be made if criteria are met to proceed to Part B: RCT. The protocol amendment will include available data from Part A and any changes to study design and statistical plan, if needed.

1.1 Study objectives

1.1.1 Primary objectives

Primary objective:	Outcome measure:
To assess the efficacy of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy in terms of ORR	ORR in all patients using Investigator assessments according to RECIST 1.1[a,b]

[a] All images will be collected and stored for possible future central re-analysis.
[b] Sensitivity analyses of ORR will be performed based on Investigator assessment according to RECIST 1.1 modified for confirmation of progression.

1.1.2 Secondary objectives

Secondary objective:	Outcome measure:
To further assess the efficacy of the combination of MEDI4736 and tremelimumab and MEDI4736 alone in terms of DoR, DCR, PFS, PFS3, PFS6, OS, OS6, and OS12	DoR, DCR, PFS, PFS3, and PFS6 in all patients using Investigator assessments according to RECIST 1.1[a] OS, OS6, and OS12
To assess the health-related QoL in metastatic PDAC patients treated with MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy as measured by EORTC QLQ-C30 global QoL scale (Part B only)	Adjusted mean change from baseline in global QoL score from the EORTC QLQ-C30 questionnaire
To assess the disease-related symptoms in metastatic PDAC patients treated with MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy as measured by EORTC QLQ-PAN26 pancreatic pain scale (Part B only)	Adjusted mean change from baseline in pancreatic pain score from the EORTC QLQ-PAN26 questionnaire
To investigate the immunogenicity of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy	Presence of ADAs for MEDI4736 and tremelimumab (confirmatory results: positive or negative; titers)
To assess the PK of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy	Concentration of MEDI4736/tremelimumab in blood and noncompartmental PK parameters (such as peak concentration and trough, as data allow; sparse sampling only)

[a] All images will be collected and stored for possible future central re-analysis.

1.1.3 Safety objectives

Safety objective:	Outcome measure:
To assess the safety and tolerability profile of MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy	AEs, physical examinations, laboratory findings (including clinical chemistry, hematology, and urinalysis), vital signs (including blood pressure and pulse), ECGs

1.1.4 Exploratory objectives

Exploratory objectives:	Outcome measure:
To assess the efficacy of MEDI4736 and tremelimumab and MEDI4736 alone according to irRECIST[a]	ORR, DoR, DCR, PFS, PFS3, and PFS6 using BICR assessments according to irRECIST[a]
To assess tolerability directly by patient self-reporting of specific PRO-CTCAE symptoms	Pre-selected PRO-CTCAE items (not more than 30) considered relevant to patients with

(Part B only)	metastatic PDAC treated with MEDI4736 or tremelimumab or SoC
To collect blood and tissue samples for defining biological responses to MEDI4736 or MEDI4736-tremelimumab and for identifying candidate markers that may correlate with likelihood of clinical benefit	Serum sPD-L1 Circulating soluble factors (eg, cytokines and autoantibodies) miRNA/mRNA T cell and MDSC phenotyping SNP genotyping
To collect and store DNA according to each country's local and ethical procedures for future exploratory research into genes/genetic variation that may influence response (ie, distribution, safety, tolerability, and efficacy) to study treatments and/or susceptibility to disease (optional genetics component)	Correlation of polymorphisms with variation in PK, pharmacodynamics, safety, or response parameters observed in patients treated with MEDI4736 and the combination of MEDI4736 and tremelimumab, and/or susceptibility to disease
To assess change in performance status in metastatic PDAC patients treated with MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy as measured by the ECOG performance status scale	ECOG performance status
To describe and evaluate resource use associated with assigned treatments and underlying disease (Part B only)	Health resource utilization methods including HOSPAD, concomitant medications, and the AE module of the electronic case report form (eCRF)
To explore the impact of treatment and disease state on health state utility using the EuroQol 5-dimension, 5-level health state utility index (EQ-5D-5L) (Part B only)	The EQ-5D-5L will be used to derive health state utility based on patient-reported data.

Note: Exploratory objective analyses may be reported separately from the main clinical study report.
[a]This analysis will be conducted only if central review data are generated at a future timepoint.

1.2 Study design

This is a second-line Phase II study assessing MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy.

This is a Phase II, open-label, multi-center study to determine the efficacy and safety of MEDI4736 evaluated as single agent or in combination with tremelimumab in patients with metastatic pancreatic ductal adenocarcinoma (PDAC) whose disease has progressed on 5-FU-containing or gemcitabine-containing first-line chemotherapy. Initially, patients will be enrolled and randomized (1:1) to treatment with MEDI4736 monotherapy or MEDI4736 + tremelimumab, until 30 patients have been randomized to treatment in each arm (Part A: Lead-in). Randomization will be stratified according to best response to prior first-line chemotherapy (Complete response [CR], Partial Response [PR], or Stable Disease [SD])

versus no response (PD) and according to prior first-line chemotherapy, 5-FU-containing versus gemcitabine-containing. Patient recruitment and tumor assessment will be monitored on an ongoing basis. The format of Part B will be determined based on the responses seen in the first 30 patients in each arm in Part A. If both arms go to Part B: Expansion, Part B; Expansion will be randomized. If only one goes, Part B: Expansion will not be randomized. Only either Part B: Expansion or Part B: RCT will be initiated. A protocol amendment will be made if criteria are met to go to Part B: RCT. A schematic diagram of the overall study design is shown in Figure 1, and a detailed study flow cart is shown in Figure 2.

Figure 1 **Phase II, MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy - study design**

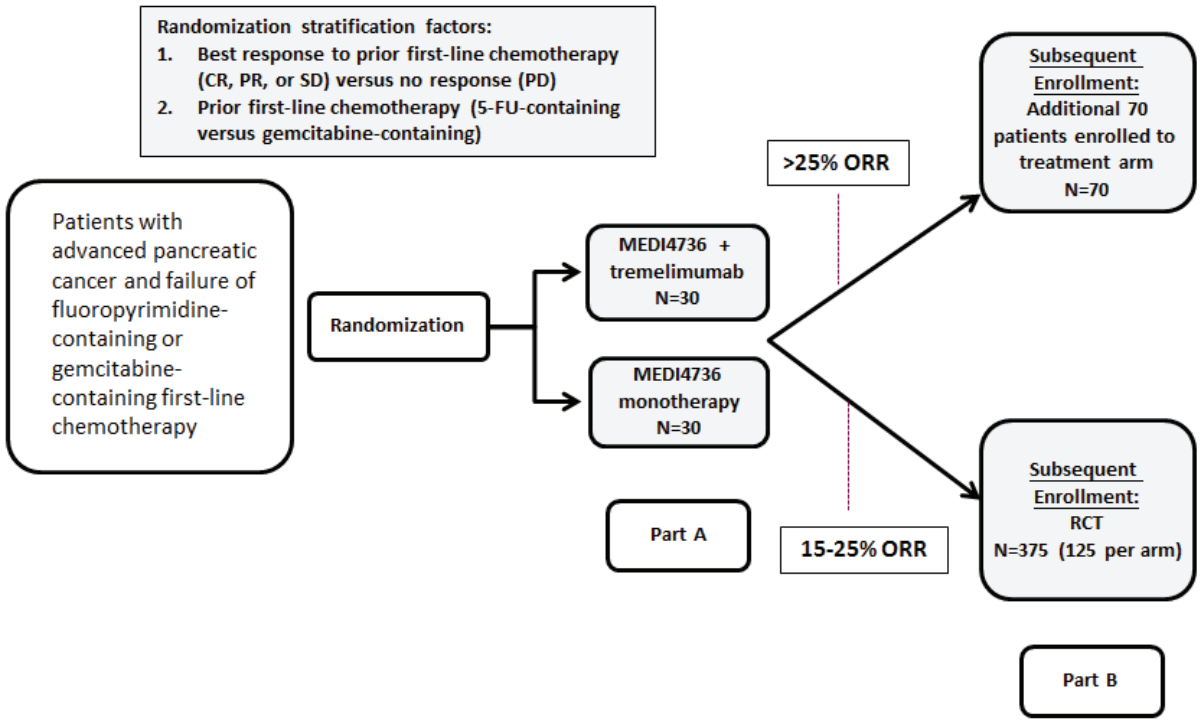
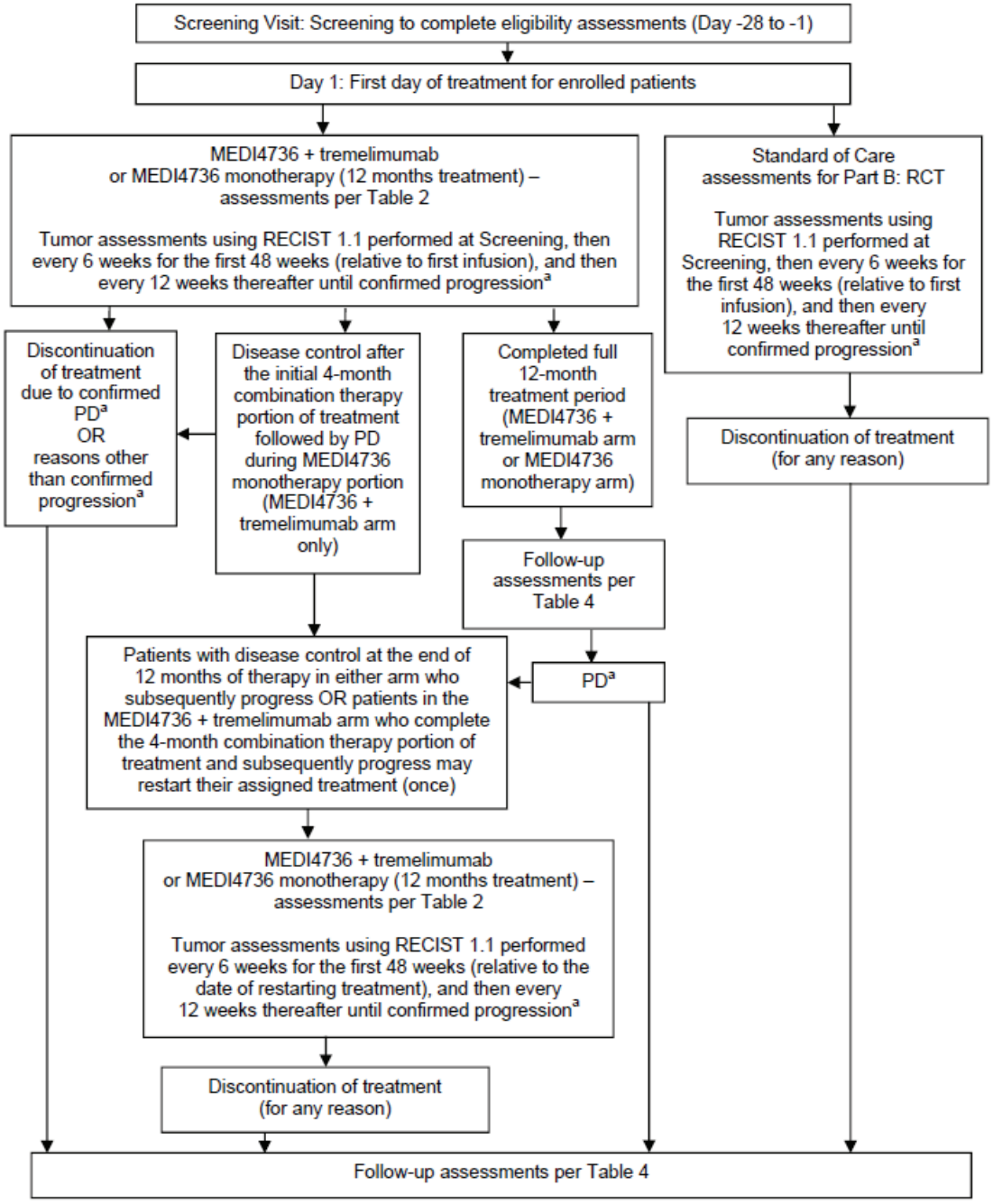


Figure 2 **Phase II, MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy - study flow chart**



1.3 Number of subjects

Part A: Lead-in and Part B: Expansion.

This study will initially enroll approximately 60 patients to receive MEDI4736 monotherapy or MEDI4736 + tremelimumab combination therapy (Part A: Lead-in). Depending on the data in each arm from Part A, an expansion phase may be initiated (Part B: Expansion approximately 140 additional patients, 70 patients per arm) and/or a randomized, controlled study (Part B: Randomized controlled trial [RCT]) may be initiated to compare MEDI4736 monotherapy and/or MEDI4736 + tremelimumab combination to Standard of Care (SoC; approximately 375 total additional patients, 125 patients per arm).

If an ORR >25% (i.e., ≥8 CR/PR) is recorded among the first 30 treated patients in either of the treatment arms in Part A (unless the totality of the data supports a different decision), an additional 70 patients will be enrolled in that arm, for a total of 100 evaluable patients per arm in the Part B: Expansion study. If both arms meet the criterion of ORR > 25%, patients will be randomised to either arm.

In Part B: Expansion, 100 patients (30 from Part A plus an additional 70 from Part B) will provide >90% power for an exact test of a single proportion, assuming a null hypothesis (H0) of ORR=10%, and an alternative hypothesis (HA) of ORR=25%, at the 2-sided 0.05 alpha level (critical value, ORR = 16.7%), while also providing adequate patients to reasonably characterize the safety profile of the expanded arm(s). A 10% response rate is considered a reasonable assumption for the average response rate achieved by current therapies in second-line advanced pancreatic cancer (Rahma et al 2013, see Section 7).

Part B: Randomized, controlled trial

Should the ORR in Part A in at least 1 arm be >15% (≥5/30 responses) and <25% (<8 responses) then a randomized controlled trial may be initiated. If initiated, this will be detailed in a separate SAP.

2. ANALYSIS SETS

2.1 Definition of analysis sets

Full Analysis Set (Intention to treat)

The Full Analysis Set (FAS) will include all randomized patients (i.e., the intent-to-treat [ITT] population) and will classify them on the basis of randomized or allocated treatment, regardless of the treatment actually received. Patients who were randomized but who did not subsequently go on to receive study treatment are included in the ITT population. For the purposes of evaluating the data in Part A to expand into Part B, only patients who actually received at least 1 dose of study treatment will be included in the FAS.

Safety Analysis Set

All patients who received at least 1 dose of IP and for whom any post-dose data are available will be included in the Safety Analysis Set, according to the treatment they actually received. When assessing safety and tolerability, summaries will be produced based on the Safety Analysis Set.

PK Analysis Set

All patients who received at least 1 dose of IP per protocol for whom any post-dose data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the PK Analysis Set. The population will be defined by the Study Physician, Pharmacokineticist, and Statistician prior to any analyses.

Definitions of the analysis sets for each outcome variable are provided in **Table 1**.

Table 1 Summary of outcome variables and analysis populations

Outcome variable	Population
Efficacy data	
PFS	FAS (ITT population)
DoR, DCR, PFS, PFS3, PFS6, OS, OS6, OS12, PROs and symptoms endpoints (*)	FAS (ITT population)
Demography	FAS (ITT population)
PK data	PK Analysis Set
Safety data	
Exposure	Safety Analysis Set
AEs	Safety Analysis Set
Laboratory measurements	Safety Analysis Set
ECOG performance status	Safety Analysis Set
Vital signs	Safety Analysis Set

(*) See Section 6

2.2 Deviations

The following general categories will be considered important deviations and be listed and discussed in the CSR as appropriate for the study.

- Deviation 1: Patients randomized (*) but who did not receive study treatment.
- Deviation 2: Patients randomized who received treatment other than that to which they were randomized.

- Deviation 3: Patients who deviate from any of the key entry criteria as per the CSP. These are inclusion criteria 3, 4, 6, 7 and 8 and exclusion criteria 4, 11 and 20.

- Deviation 4: Baseline RECIST scan > 28 days before start of randomized treatment.

- Deviation 5: No baseline RECIST 1.1 assessment on or before date of first dose.

- Deviation 6: Received prohibited concomitant medications (including other anti-cancer agents). Please refer to the CSP Section 7.7. This will be used as a guiding principle for the physician review of all medications prior to database lock to identify those likely have an impact on efficacy.

(*): where applicable

In addition to the programmatic determination of the deviations above (based on the clinical database), monitoring notes or summaries will be reviewed to determine any important post entry deviations that are not identifiable via programming, and to check that those identified via programming are correctly classified. The final classification of deviations will be made at the data review meeting (DRM) prior to database lock or data freeze. Decisions made at the DRM will be documented and approved by AstraZeneca prior to analysis.

Deviation 1 will lead to exclusion from the Safety analysis set. None of the other deviations will lead patients being excluded from the analysis sets described in Section 2.1 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK).

Errors in treatment dispensing, in addition to incorrect stratifications, will also be summarized and listed separately from the important protocol deviations. If a patient is not randomized or treated according to the randomization schedule, it is envisaged that there will be 2 subcategories:

- Patients who receive no treatment whatsoever for a period of time due to errors in dispensing of medication. Note, this is not due to tolerability issues where patients may stop taking drug.

- The patient receives a treatment pack with a different code to their randomization code. However, the actual treatment may still match the randomized treatment. For example, a patient is given randomization code 0001, which according to the randomization schedule is MEDI4736. However, at the randomization visit they are given treatment pack 0003, but this still contains MEDI4736.

The summary will include all patients with a dispensing error but will also include information on how many of those patients received at least one dose of the wrong treatment at any time. Patients who receive the wrong treatment at any time will be included in the Safety Analysis Set as described in Section 2.1. During the study, decisions on how to handle misrandomizations will be made on an individual basis with written instruction from the study team leader and/or statistician.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Response

For all subjects, the RECIST tumor response data will be used to determine each subject's visit response according to RECIST version 1.1. It will also be used to determine if and when a subject has progressed in accordance with RECIST and also their best objective response.

The methods of assessment of tumor burden used at baseline are computed tomography (CT) and/or magnetic resonance imaging (MRI) scans of the chest and abdomen; scans of the pelvis should only be done when there is suspected or documented disease involvement. Any other areas of disease involvement should be additionally imaged based on the signs and symptoms of individual patients.

The RECIST 1.1 guidelines for measurable, nonmeasurable, target, and non-target lesions and the objective tumor response criteria (CR, PR, SD, or PD) are presented in CSP Appendix F.

Baseline radiological tumor assessments are to be performed no more than 28 days before the start of randomized treatment and ideally as close as possible to the start of study treatment. Tumor assessments are then performed every 6 weeks (q6w±7 days) for the first 48 weeks relative to the date of the first infusion, and q12w±7 days thereafter until confirmed objective disease progression.

If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at his/her scheduled visits. This schedule is to be followed in order to minimize any unintentional bias caused by some patients being assessed at a different frequency than other patients.

Categorization of objective tumor response assessment will be based on the RECIST 1.1 criteria of response: CR, PR, SD, and PD. At each visit, patients will be programmatically assigned a response based on tumor assessments provided by the investigator for target lesions, non-target lesions and new lesions. Target lesion progression will be calculated in comparison to when the tumor burden was at a minimum (i.e., smallest sum of diameters previously recorded on study). In the absence of progression, tumor response (CR or PR) and SD will be calculated in comparison to the baseline tumor measurements obtained before starting treatment. If a patient has had a tumor assessment that cannot be evaluated, then the patient will be assigned a visit response of not evaluable (NE; unless there is evidence of progression, in which case the response will be assigned as PD). Endpoints (of ORR, DoR, DCR, PFS, PFS3, and PFS6) will be derived from the overall visit response date and the scan dates.

Note that overall responses as per Investigator opinion will not be used for any endpoint based on RECIST criteria.

A confirmatory scan is required following the initial demonstration of PD. The confirmatory scan should occur preferably at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinically significant deterioration. Treatment will continue between the initial assessment of progression and confirmation for progression.

If a patient discontinues treatment (and/or receives a subsequent anticancer therapy) prior to progression, then the patient should still continue to be followed until confirmed objective disease progression.

Objective tumor response (CR or PR) should be confirmed preferably at the next scheduled visit and not less than 4 weeks after the visit when the response was first observed.

3.1.1 Target lesions (TLs)

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with CT or MRI and which is suitable for accurate repeated measurements.

A subject can have a maximum of 5 measurable lesions recorded at baseline with a maximum of 2 lesions per organ (representative of all lesions involved suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest and prior to randomization will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as non-target lesions (NTLs) at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

TL response will be derived based on the Investigator's overall assessment of TLs as follows (see **Table 2**):

Table 2 TL visit responses

Visit Responses	Descriptions
Complete Response (CR)	Disappearance of all TLs since baseline. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters.
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.
Progressive disease (PD)	At least a 20% increase in the sum of diameters of TLs taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5%.
Not Evaluable (NE)	Only relevant if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit . Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response.

CR Complete response; PR Partial response; PD Progression of disease; NE Not evaluable; SD Stable disease; TL Target lesion.

Rounding of TL data

For calculation of PD and PR for TLs percentage changes from baseline and previous minimum should be rounded to 1 decimal place (d.p.) before assigning a TL response. For example 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%

Missing TL data

For a visit to be evaluable all TL measurements should be recorded. However, a visit response of PD should still be assigned if any of the following occurred:

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

Note: the nadir can only be taken from assessments where all the TLs had a LD recorded.

If there is at least one TL measurement missing and a visit response of PD cannot be assigned, the visit response is not evaluable (NE).

Lymph nodes

For lymph nodes, if the size reduces to < 10mm then these are considered non-pathological. However a size will still be given and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are < 10mm and all other TLs are 0mm then although the sum may be >0mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

A CR can only be followed by CR, PD or NE. If a CR has occurred then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node LD increases by 20% but remains < 10mm.

- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to NE irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.

- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD

- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size ('x') above which it cannot be accurately measured should be recorded. If using a value of x in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team blinded to treatment assignment. It is expected that a visit response of PD will remain in the vast majority of cases.

TL too small to measure

If a TL becomes too small to measure a value of 5mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and entered a smaller value that can be reliably measured. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

Previously irradiated lesions (i.e. lesion irradiated prior to entry into the study) should be recorded as NTLs and should not form part of the TL assessment.

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolization), should be handled in the following way and once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumors:

- Step 1: The diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and if ≤ 1/3 of the TLs have missing measurements then scale up as described in the 'Scaling' section below. If the scaling results in a visit response of PD then the subject would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then, if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs have missing measurements), treating the lesion with intervention as missing, and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subject to intervention also have a value of 0 recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

At subsequent visits, the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with intervention should be treated as missing and scaled up (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention)

If > 1/3 of TL measurements are missing (because of intervention) then TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e. if using a value of 0 for missing lesions, the sum of diameters has still increased by 20% or more compared to nadir and the sum of TLs has increased by ≥5mm from nadir).

If ≤ 1/3 of the TL measurements are missing (because of intervention) then the results will be scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters and this will be used in calculations; this is equivalent to comparing the visit sum of diameters of the non-missing lesions to the nadir sum of diameters excluding the lesions with missing measurements).

Example of scaling

Lesion	Longest diameter at nadir visit (mm)	Longest diameter at follow-up visit (mm)
1	7.2	7.1
2	6.7	6.4
3	4.3	4.0
4	8.6	8.5
5	2.5	Intervention
Sum	**29.3**	**26**

Lesion 5 is missing at the follow-up visit.
The sum of lesions 1-4 at the follow-up is 26 mm. The sum of the corresponding lesions at nadir visit is 26.8 mm.

Scale up as follows to give an estimated TL sum of 28.4mm:

$$\frac{26}{26.8} \times 29.3 = 28.4mm$$

CR will not be allowed as a TL response for visits where there is missing data. Only PR, SD or PD (or NE) could be assigned as the TL visit response in these cases. However, for visits with ≤1/3 lesion assessments not recorded, the scaled up sum of TLs diameters will be included when defining the nadir value for the assessment of progression.

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions should be summed and reported as the LD for the lesion that split.

Lesions that merge

If two TLs merge, then the LD of the merged lesion should be recorded for one of the TL sizes and the other TL size should be recorded as 0cm.

Change in method of assessment of TLs

CT, MRI and clinical examination are the only methods of assessment that can be used within a trial, with CT and MRI being the preferred methods and clinical examination only used in special cases. If a change in method of assessment occurs between CT and MRI this will be considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions should be treated as missing.

3.1.2 Non-Target Lesions (NTLs) and new lesions

At each visit an overall assessment of the NTL response should be recorded by the investigator. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the Investigator's overall assessment of NTLs as follows (see **Table 3**):

Table 3 NTL visit Responses

Visit Response	Description
Complete Response (CR)	Disappearance of all NTLs since baseline. All lymph nodes must be non-pathological in size (<10 mm short axis).
Non-CR/Non-PD	Persistence of one or more NTLs.
Progressive Disease (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Not Evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the Investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.

CR Complete response; PR Partial response; PD Progression of disease; NE Not evaluable; NTL Non-target lesion; TL Target lesion.

To achieve 'unequivocal progression' on the basis of NTLs, there must be an overall level of substantial worsening in non-target disease such that, even in the presence of SD or PR in TLs, the overall tumor burden has increased sufficiently to merit a determination of disease progression. A modest 'increase' in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.
Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

The finding of a new lesion should be unequivocal: i.e. not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumor.

New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.

If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present, but should not overtly affect the derivation.

Symptomatic progression is not a descriptor for progression of NTLs: it is a reason for stopping study therapy and will not be included in any assessment of NTLs.

Patients with 'symptomatic progression' requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumor assessments where possible until objective disease progression is observed.

3.1.3 Overall visit response

Table 4 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 4 Overall visit responses

TARGET	NON-TARGET	NEW LESIONS	OVERALL VISIT RESPONSE
CR	CR	No	CR
CR	NA	No	CR
CR	Non-CR/Non-PD	No	PR
CR	NE	No	PR
PR	Non-PD or NE	No	PR
SD	Non-PD or NE	No	SD
NE	Non-PD or NE	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR Complete response, PR Partial response, SD Stable disease, PD Progression of disease, NE Not evaluable, NA Not applicable (only relevant if there were no non-target lesions at baseline)

3.1.4 Independent Review (Not applicable)

3.2 Outcome variables

3.2.1 Primary endpoint - Objective Response Rate (ORR)

The primary endpoint of ORR, according to investigator assessment, is defined as the number (%) of subjects with at least one visit response of confirmed CR or PR and will be based on the FAS.

A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging, preferably at the next regularly scheduled imaging visit and not less than 4 weeks after the visit when the response was first observed with no evidence of progression between the initial and CR/PR confirmation visits. Therefore, data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue randomized treatment without progression, receive a subsequent anti-cancer therapy and then respond will not be included as responders in the ORR.

For **sensitivity analysis**, ORR will be assessed using the RECIST Investigator assessed tumor data following a modification where any objective progression requires confirmation. Therefore, data obtained up until confirmed progression, or last evaluable assessment in the absence of a confirmed progression, will be included in the assessment of ORR. Note that the response may be after an unconfirmed progression.

3.2.2 Secondary endpoints

3.2.2.1 Progression Free Survival

Progression free survival (PFS) is defined as the time from the date of randomization (for Part A, and Part B if randomization is applicable) or first dosing (for Part B if randomization is not applicable)until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the subject withdraws from randomized therapy or receives another anti-cancer therapy prior to progression.
Subjects who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last evaluable RECIST 1.1 assessment. However, if the subject progresses or dies after two or more missed visits, the subject will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the two missed visits.

Given the scheduled visit assessment scheme (i.e. six-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 288 (i.e. Week 41) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e., 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to twelve-weekly this will equate to 20 weeks (i.e., take the average of 6 and 12 weeks which gives 9 weeks and then apply same rationale hence 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks). The time period for the previous RECIST assessment will be from study days 288 to 344 (i.e. Week 41 to Week 49). From Week 49 onwards (when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 26 weeks (i.e. 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks).
If a patient has no evaluable visit or does not have baseline data, he/she will be censored at date of randomization unless they die within 2 visits from randomization date (2 x 6 weeks + 1 week for a late assessment).

If the subject has no post-baseline evaluable visit or does not have baseline data they will be censored at date of first dose unless they die within 2 visits of first dose (12 weeks plus 1 week allowing for a late assessment within the visit window).

The PFS time will always be derived based on scan/assessment dates not visit dates. RECIST assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- For investigational assessments, the date of progression will be determined based on the earliest of the dates of the component that triggered the progression
- When censoring a subject for PFS the subject will be censored at the latest of the dates contributing to a particular overall visit assessment.

Note: For TLs, only the latest scan date is recorded out of all scans performed at that assessment for the TLs and similarly for NTLs only the latest scan date is recorded out of all scans performed at that assessment for the NTLs.

Time to study discontinuation or Death

Time to Study Discontinuation or Death is defined as the time from first dosing to the earlier of the date of study treatment discontinuation or death. Any patient not known to have died at the time of analysis and not known to have discontinued study treatment will be censored based on the last recorded date on which the patient was known to be alive.

3.2.2.2 Proportion of patients alive and progression-free

The proportion of patients alive and progression free at 3 months (i.e., PF3) will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by the site investigator) at 3 months.

The proportion of patients alive and progression free at 6 months (i.e., PF6) will be defined as the Kaplan-Meier estimate of PFS (per RECIST 1.1 as assessed by the site investigator) at 6 months.

3.2.2.3 Overall Survival

Overall survival (OS) is defined as the time from the date of randomization (for Part A, and Part B if randomization is applicable) or first dosing (for Part B if randomization is not applicable) until death due to any cause (i.e. date of death or censoring – date of first dosing or randomization + 1). Any subject not known to have died at the time of analysis will be censored based on the last recorded date on which the subject was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Note: Survival calls will be made in the week following the date of data cut-off (DCO) for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from

the study) and "lost to follow-up" patients at the time of the final OS analysis should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly-available death registries. In the event that the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly-available resources where it is possible to do so under applicable local laws.

Note that for any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since the SURVIVE module is only completed for patients off treatment).

3.2.2.4 Proportion of patients alive

The proportion of patients alive at 6 months (OS6) will be defined as the Kaplan-Meier estimate of OS at 6 months.

The proportion of patients alive at 12 months (OS12) will be defined as the Kaplan-Meier estimate of OS at 12 months.

3.2.2.5 Duration of Response

Duration of response (DoR) will be defined as the time from the date of first documented response until date of documented progression or death in the absence of disease progression, the end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of PR or CR.

If a subject does not progress following a response, then their duration of response will use the PFS censoring time. DoR will not be defined for those patients who do not have documented response.

Additionally, DoR will be obtained using the algorithm described (see Appendix F in the CSP), but following a modification where any objective progression requires confirmation. Therefore, the end of response should coincide with the date of progression or death for any cause. Note that the time of initial response may be after an unconfirmed progression.

3.2.2.6 Best objective response

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in Appendix F in the CSP. It is the best response a patient has had following date of first dosing but prior to starting any subsequent cancer therapy and prior to RECIST progression or the last evaluable assessment in the absence of RECIST progression.
Categorization of BoR will be based on RECIST using the following response categories: CR, PR, SD, PD, and NE.

CR or PR must be confirmed. For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 6 weeks minus 1 week, i.e. at least 35 days (to allow for the assessment window), after the first dosing. For CR/PR, the initial overall visit assessment which showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

BoR will be determined programmatically based on RECIST from the overall visit response using all Investigator assessment data up until the first progression event. The denominator will be consistent with the one used in the ORR analysis. Furthermore, it will be determined programmatically based on RECIST modified for confirmation of progression. This will use all data up until the progression event that is used for the analysis (i.e., unconfirmed progressions are not considered progression events, which means that the BoR may be after an unconfirmed progression for some patients).

For patients whose progression event is death, BoR will be calculated based upon all evaluable RECIST assessments prior to death.

For patients who die with no evaluable RECIST assessments, if the death occurs ≤13 weeks (i.e., 2*6 weeks + 1 week to allow for a late assessment within the assessment window) after the first dosing, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >7 weeks (i.e., 6 weeks + 1 week) after the first dosing then BoR will be assigned to the NE category.

Progression events that have been censored because they occurred more than two missed visits after the last evaluable assessment will not contribute to the BoR derivation.

3.2.2.7 Change in tumor size

For supportive purposes percentage change from baseline in tumor size will be derived at each scheduled tumor assessment visit (i.e., Week 6, Week 12 etc… hereafter referred to as Week X for convenience). Best percentage change from baseline in tumor size will also be derived as the biggest decrease or the smallest increase in tumor size from baseline.

This is based on RECIST target lesion measurements taken at baseline and at the timepoint of interest. Tumor size is defined as the sum of the longest diameters of the target lesions. Target lesions are measurable tumor lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to starting treatment (see Section 3.1 for more details). The change in target lesion tumor size at Week X will be obtained for each patient by taking the difference between the sum of the target lesions at Week X and the sum of the target lesions at baseline. To obtain the percentage change in target lesion tumor size at Week X the change in target lesion tumor size is divided by the sum of the target lesions at baseline and multiplied by 100 (i.e. (Week X - baseline) / baseline * 100). More details on target lesions and measurements can be found in Section 3.1.

Apply a window around the Week X visit: Whenever tumor size data for the Week X visit (Note: or visit at which progression was documented if before Week X) is available then this should be used in the analysis. A windowing rule will be applied and will follow the protocol allowed visit window; therefore any RECIST scan performed within ± 7 days of the protocol scheduled visit will be used for that visit.

The above derivations will be programmed for the investigator assessments based upon RECIST analysis.

3.2.2.8 Disease Control Rate (DCR)

Disease control rate at 3 months is defined as the percentage of patients who have a BoR of CR or PR in the first 3 months (i.e. 12+1=13 weeks to allow for a late assessment within the assessment window) or who have demonstrated SD for a minimum interval of 12 weeks (minus 1 week to allow for an early assessment within the assessment window, i.e., 77 days) following the start of treatment. Disease control rate at 6 months is defined as the percentage of patients who have a BoR of CR or PR in the first 6 months (i.e. 24+1=25 weeks to allow for a late assessment within the assessment window) or who have demonstrated SD for a minimum interval of 24 weeks (minus 1 week to allow for an early assessment within the assessment window, i.e., 161 days) following the start of treatment. DCR at 12 months is defined as the percentage of patients who have a BoR of CR or PR in the first 12 months (i.e. 48+1=49 weeks to allow for a late assessment within the assessment window) or who have demonstrated SD for a minimum interval of 48 weeks (minus 1 week to allow for an early assessment within the assessment window, i.e., 329 days) following the start of treatment.

DCR will be determined programmatically based on RECIST 1.1 using Investigator assessments and all data up until the first progression event. This will use all data up until the progression event that is used for the analysis.

3.3 Safety

Safety and tolerability will be assessed in terms of adverse events (AEs) (including serious adverse events [SAEs]), deaths, laboratory data, vital signs, electrocardiograms (ECGs) and exposure. These will be collected for all patients.

Data from the initial treatment period (i.e., the initial 12 months of treatment) and safety data from the retreatment period may also be summarized separately (see Section 4.1). ‘On treatment’ will be defined as assessments between date of start dose and 90 days following last dose of the study treatment on each period of treatment. The Safety Analysis Set will be used for reporting of safety data.

3.3.1 Adverse events (AE)

AEs and SAEs will be collected throughout the study from the time the informed consent is signed through 90 days after the last dose of the study treatment.

The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 4.03). For those events without assigned CTCAE grades, the recommendation in the CTCAE criteria that converts mild, moderate, and severe events into CTCAE grades should be used.

A treatment emergent adverse event (TEAE) is an AE with an onset date or a pre-existing AE worsening following the first dose of study treatment through to 90 days after the last dose of the study treatment. In the unlikely event of the components being administered separately then date of first dose/last dose will be considered as the earliest/latest dosing date of either component.

Other significant adverse events

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and AEs leading to discontinuation. Based on the expert's judgment, significant AEs of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered other significant adverse events (OAEs) and reported as such in the CSR. A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

Examples of these are marked hematological and other laboratory abnormalities, and certain events that lead to intervention (other than those already classified as serious) or significant additional treatment.

AEs of special interest

Adverse events of special interest (AESIs) are events of scientific and medical interest specific to the further understanding of the MEDI4736 and tremelimumab safety profile and require close monitoring and rapid communication by the Investigator to AstraZeneca. MEDI4736 and tremelimumab AESIs may be serious or non-serious. The rapid reporting of these AESIs allows ongoing analysis of these events in order to characterize and understand them in association with the use of these IPs.

AESIs for MEDI4736 include but are not limited to events with a potential inflammatory or immune-mediated mechanism and which may require more frequent monitoring and/or interventions such as steroids, immunosuppressants and/or hormone replacement therapy. These AESIs are being closely monitored in clinical studies with MEDI4736 monotherapy and combination therapy. An immune-related adverse event (irAE) is defined as an adverse event that is associated with drug exposure and is consistent with an immune-mediated mechanism of action and where there is no clear alternate aetiology. Serologic, immunologic, and histologic (biopsy) data, as appropriate, should be used to support an irAE diagnosis. Appropriate efforts should be made to rule out neoplastic, infectious, metabolic, toxin, or other etiologic causes of the irAE.

Some clinical concepts (including some selected individual preferred terms and higher level terms) have been considered "AEs of special interest" to the MEDI4736 program. These AESIs have been identified as Pneumonitis, Colitis, Select hepatic events, Hypothyroidism, Hyperthyroidism, Hypophysitis, Adrenal Insufficiency, Dermatitis, Select renal events, Select pancreatic events, and Infusion related/Hypersensitivity/Anaphylactic reactions, Diarrhea, and Rash. Other categories may be added or existing terms may be merged as necessary. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AEs of interest and identified which preferred terms contribute to each AESI. A further review will take place prior to Database lock (DBL) to ensure any further terms not already included are captured within the categories.

3.3.2 Treatment exposure

Exposure will be defined separately for the initial treatment period and for the re-treatment period as follows:

Total (or intended) exposure of study treatment

- Total (or intended) exposure = last dose date +27 days where dose > 0 mg or death or DCO, whichever occurs first – first dose date + 1

Actual exposure of study treatment

- Actual exposure = intended exposure – total duration of dose delays, where intended exposure will be calculated as above.

A dose delay is stated as soon as the infusion is not administered on the day planned per dosing scheduling.

Dose reductions are not permitted per Section 6.7 of the CSP. The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Exposure will also be measured by the number of infusions received.

Patients who permanently discontinue during a dose delay

If a decision is made to permanently discontinue study treatment in-between cycles or during a dose delay then the date of last administration of study medication recorded will be used in the programming.

3.3.3 Dose intensity

For each molecule, standard dose intensity will be derived separately for the initial treatment period.

Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation.

Relative dose intensity (RDI) will be defined as follows:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing.

As the treatment period is for a finite period of time, additional considerations need to be taken into account for the above calculations. In the absence of event by the end of the initial treatment period, the following censoring will be applied, relative to the first infusion in that period: earliest between the date of censoring used in the PFS and

- MEDI4736: end of period = 48 weeks (+1 week for late assessment)
- Tremelimumab: end of period = 12 weeks (+1week for late assessment)

The standard dose intensity calculations above are focused upon standard agents that treat to progression and as the primary analysis is based upon RECIST 1.1 it is still appropriate to calculate these. However, due to the fact that there is the intention to treat through progression per the protocol there will be an additional measure of dose intensity to investigate the entire treatment period including the time period after progression:

- Relative dose intensity (treatment through progression) (RDI2) = 100% * d/D, where d is the actual cumulative dose delivered up to the later of progression (or a censoring event) or the actual last day of dosing and D is the intended cumulative dose up to the later of progression (or a censoring event) or the actual last day of dosing.

The censoring event mentioned above is the date of censoring used in the PFS analysis.

3.3.4 Time to first subsequent therapy from discontinuation of study treatment

Time to first subsequent therapy from date of last dose is defined as the time from the date of discontinuation of study treatment to the start date of the first subsequent therapy. Any palliative radiotherapy post discontinuation of study treatment should not be considered as subsequent cancer therapy. Any patient not known to have had a first subsequent therapy will not have this calculation performed.

3.3.5 Laboratory data

Laboratory data will be collected throughout the study, from screening to the follow-up visits as described in the CSP. Blood and urine samples for determination of hematology, clinical chemistry, and urinalysis will be collected as described in Section 5.2.1 of the CSP. For the definition of baseline and the derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 3.3.8 below will be used.

Change from baseline in hematology and clinical chemistry variables will be calculated for each post-dose visit on treatment. CTC grades will be defined at each visit according to the CTC grade criteria using local or project ranges as required, after conversion of lab result to corresponding preferred units. The following parameters have CTC grades defined for both high and low values: potassium, sodium, magnesium, glucose and corrected calcium so high and low CTC grades will be calculated.

Corrected calcium will be derived during creation of the reporting database using the following formulas:

Corrected calcium (mmol/L) = Total calcium (mmol/L) + ([40 – albumin (G/L)] x 0.02)

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range) and high (above range).

The maximum or minimum on-treatment value (depending on the direction of an adverse effect) will be defined for each laboratory parameter as the maximum (or minimum) post-dose value at any time.

Project reference ranges will be used throughout for reporting purposes. The denominator used in laboratory summaries of CTC grades will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality.

For example:

- If a CTCAE criterion involves a change from baseline, evaluable patients would have both a pre-dose and at least 1 post-dose value recorded

- If a CTCAE criterion does not consider changes from baseline, to be evaluable the patient need only have 1 post dose-value recorded.

3.3.6 ECGs

ECG data are collected at screening and as clinically indicated while on study treatment. For derivation of post baseline visit values considering visit window and to handle multiple records present in any visit window, derivation rules as described in Section 3.3.8 below will be used.

At each time point the Investigator's assessment of the ECG will be collected locally. Heart rate, duration of QRS complex, RR, PR and QT intervals will be collected centrally via a digital read.

For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point.

3.3.7 Vital signs

Vital signs data obtained up until the 30 days (+3 days as per permitted time-window) from date of last dose of study treatment will be used for reporting. For derivation of post baseline visit values considering visit window and to handle multiple records present in any visit window, derivation rules as described in Section 3.3.8 below will be used.

3.3.8 General considerations for safety assessment

Time windows will be defined for any presentations that summarize values by visit. The following conventions will apply:

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarized. Inclusion within the time window should be based on the actual date and not the intended date of the visit.

- All unscheduled visit data should have the potential to be included in the summaries.

  The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.

  For example, the visit windows for vital signs data for MEDI4736 + tremelimumab combination therapy and MEDI4736 monotherapy (with 4 weeks between scheduled assessments) are:

  (Week 4) Day 29, visit window 2 – 42

  (Week 8) Day 57, visit window 43 – 70

  (Week 12) Day 85, visit window 71 – 98

  (Week 16) Day 113, visit window 99 – 126

  (Week 20) Day 141, visit window 127 – 154

  (Week 24) Day 169, visit window 155 – 182

  (Week 28) Day 197, visit window 183 – 210

  (Week 32) Day 225, visit window 211 – 238

  (Week 36) Day 253, visit window 239 – 266

  (Week 40) Day 281, visit window 267 – 294

(Week 44) Day 309, visit window 295 – 322

(Week 48) Day 337, visit window 323 – 350

Note: Due to the differing assessment schedules the visit windows will be different for the different study treatments and endpoints.

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).

- Listings should display all values contributing to a time point for a patient.

- For visit- based summaries:

  – If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be summarized, or the earlier in the event the values are equidistant from the nominal visit date. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all post baseline values collected are used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date.

  – To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group visit data should only be summarized if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.

- For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit based summary, when deriving a patient level statistic such as a maximum.

- Baseline will be defined as the last non-missing measurement prior to dosing with study treatment. For the re-treatment period, baseline is similarly defined as the last non-missing measurement prior to the first dose on the re-treatment period. For laboratory data and height/weight assessments, any assessments made on Day 1 will be considered pre-dose. Alternatively, if two visits are equally eligible to assess patient status at baseline (e.g., screening and baseline assessments both on the same date prior to first dose with no washout or other intervention in the screening period), the average will be taken as a baseline value. For non-numeric laboratory tests (i.e. some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline. Where safety data are summarized over time, study day will be calculated in relation to date of first treatment. ECOG performed on Day 1 will be considered as baseline.

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings. Additionally, adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

3.4 Biomarker variables

PD-L1 expression status (positive, negative) is defined in **Table 5**.

Table 5 PD-L1 status definition

Interpretation	Staining description
Positive for PD-L1	≥25% tumor cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control OR ≥25% tumor associated immune cell positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control
Negative for PD-L1	<25% tumor cell membrane positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control AND <25% tumor associated immune cell positivity for PD-L1 at any intensity above background staining as noted on the corresponding negative control

Abbr: PD-L1 Programmed cell death ligand 1

3.5 Pharmacokinetic and Immunogenicity variables

Samples for determination of MEDI4736 and tremelimumab concentration in serum will be analyzed by a designated third party laboratory on behalf of AstraZeneca.

3.5.1 Population pharmacokinetics and exposure-response/safety analysis

A population PK model analysis might be performed using a nonlinear mixed-effects modelling approach. The impact of physiologically-relevant patient characteristics (covariates) and disease on PK will be evaluated. The relationship between the PK exposure and the effect on safety and efficacy endpoints will be evaluated. The results of such an analysis will be reported in a separate report. The PK, pharmacodynamics (PDx), demographic, safety, and efficacy data collected in this study may also be combined with similar data from other studies and explored using population PK and/or PK-PDx methods. The mean PK profiles will be will be shown in a figure.

3.5.2 Pharmacokinetic non-compartmental analysis

PK concentration data and summary statistics will be tabulated by nominal visit/time point. Individual and mean blood concentration-time profiles will be generated. Samples below the lower limit of quantification will be treated as missing in the analyses.

3.5.3 Immunogenicity analysis

Immunogenicity results will be analyzed descriptively by summarizing the number and percentage of patients who develop detectable ADAs against MEDI4736 monotherapy or MEDI4736 in combination with tremelimumab. The immunogenicity titer and presence of neutralizing ADAs will be reported for samples confirmed positive for the presence of ADAs. The effect of immunogenicity on PK, pharmacodynamics, efficacy, and safety will be evaluated, if the data allow, but will be reported in a separate report.

3.6 Health resource Use (Not applicable)

3.7 Pharmacogenetics

A statistical analysis plan will be prepared where appropriate and reported outside the CSR.

4. ANALYSIS METHODS

4.1 General principles

The below mentioned general principles will be followed throughout the study:

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarised by the number of observations, mean, standard deviation, median, minimum, and maximum. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables will be summarised by frequency counts and percentages for each category.

- Unless otherwise stated, percentages will be calculated out of the population total and for each treatment group.

- For continuous data the mean and median will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.

- For categorical data, percentages will be rounded to 1 decimal place.

- SAS® version 9.2 or above will be used for all analyses.

- **All analyses related to Part B described in this document are optional, unless this study really recruits patients for Part B.**

Baseline will be the last assessment of the variable under consideration prior to the intake of the first dose of investigational product, except for efficacy variables. For efficacy variables, baseline is defined as the last visit prior to treatment start date.

Concomitant medications will be medications taken at least once from the first dosing date until 90 days after the last dose of the study treatment. Medications with missing start or end dates will also be assumed to be concomitant (whatever is the ongoing status).

When the date of birth is incomplete due to local regulatory constraints, the age as collected in the eCRF will be used for analysis purpose.

Handling of missing/incomplete dates

The original incomplete or missing dates will be presented in the listings.

- Adverse events: all AEs will be considered as treatment-emergent unless the opposite can be clearly stated. Imputation will be done only in the context of identifying TEAEs.

- Concomitant medications: all medications will be considered as concomitant unless the opposite can be clearly stated.

No other imputation will be made.

The following other general principles will also apply:

- All data collected will be listed.

- Efficacy data will be summarized and analyzed based on the FAS. PK data will be summarized and analyzed based on the PK Analysis Set. Safety and treatment exposure data will be summarized based upon the Safety Analysis Set. Study population and demography data will be summarized based upon the FAS.

- All outputs will be summarized by treatment arm for all enrolled patients (ITT).

- If any arm goes to Part B: Expansion, safety data collected during Part A and Part B: Expansion will be presented once, at the time of Part B: Expansion analysis. The decision for Part B is to be made on efficacy data, these will be presented at the time of Part A analysis (see Section 5.2) and thereafter pooled with Part B: Expansion data if any arm goes to Part B: Expansion. If any arm goes to Part B: Expansion, data collected during Part B: Expansion will not be presented separately.

- Safety data will be summarized from the initial treatment period (i.e. the initial 12 months of treatment). Safety data from the re-treatment period may also be summarized via a small set of headline summaries should there be sufficient number of patients re-treated to warrant this. Any safety summaries representing the re-treatment period will be based upon a subset of the Safety Analysis Set representing patients who have had at least one dose of study treatment in the re-treatment period.

Table 6 details which endpoints are to be analyzed, together with preplanned sensitivity analyses indicating which analysis is regarded as primary for that endpoint. Formal statistical analysis will be done only for ORR for Part B: Expansion.

Unless otherwise stated, data and endpoints derived from RECIST tumor assessments will refer to Investigator assessed data.

Table 6 Preplanned statistical and sensitivity analyses to be conducted

Endpoints analyzed	Notes
Objective response rate	Summarized by treatment arm with lower 95% Clopper-Pearson confidence limits (Part A) Part B: Expansion - Formal analysis using exact test for a binomial proportion, with 2-sided 95% Clopper-Pearson CI
Duration of response	Kaplan Meier plots and summaries
DCR	Summarized by treatment arm n (%).
Best objective response	Summary statistics, N (%)
Progression free survival, PFS3, PFS6	Kaplan Meier plots, estimates and summaries as appropriate
Overall survival, OS6, OS12	Kaplan Meier plots and summaries.

4.2 Analysis methods

4.2.1 Multiplicity (Not applicable)

4.2.2 Primary endpoint – Objective Response Rate

The ORR will be based on the programmatically derived RECIST using the Investigator tumor data. An additional sensitivity analysis will be performed on programmatically derived ORR using Investigator data (RECIST modified for confirmation of progression) to determine if there is any difference when using progression confirmation rules.

Summaries will be produced that present the number and percentage of patients with a tumor response (CR/PR) with lower 95% Clopper-Pearson confidence limits for Part A and with a formal analysis using exact test for a binomial proportion, with 2-sided 95% Clopper-Pearson confidence interval for part B: Expansion. For Part A primary analysis, the number and percentage of patients with a tumor response (and associated CI) will also be presented where unconfirmed responses are no longer excluded. Overall visit response data will be listed.

For each treatment arm, best objective response (BoR) will be summarized by n (%) for each category (CR, PR, SD, PD, and NE). No formal statistical analyses are planned for BoR.

4.2.3 Duration of response

Descriptive data will be provided for the DoR in responding patients.

If data allows, Kaplan-Meier plots of the proportion of patients in responders will be presented by treatment arm.

In addition, this will be produced for DoR when obtained while objective progression requires confirmation.

4.2.4 Disease control rate

Summary statistics (i.e., number of patients [%] with 95% CIs) on DCR at 3, 6 and 12 months will be produced by treatment arm. It will be provided for both overall and by Korea status (i.e. Korea and Non Korea).

4.2.5 Progression-free survival

Kaplan-Meier plots of PFS will be presented for each arm. Summaries of the number and percentage of patients who have died, are still in survival follow-up, are lost to follow-up and have withdrawn consent will be provided along with median PFS for each cohort (if calculable). It will be provided for both overall and by Korea status (i.e. Korea and Non Korea).

Exploratory analysis

PFS based on RECIST 1.1 modified for confirmation of progression will be performed for exploratory purposes using the algorithm described in CSP Appendix F, but following a modification whereby any objective disease progression must be confirmed by the next scheduled scan (see Section 3.1). 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. Patients with a single disease progression and no further tumor assessment scans will be treated as PD in the analysis.

The same analysis as above will be performed for this exploratory analysis.

Proportion of patients alive and progression free after 3 months

The PFS rate at 3 months (PFS3) will be calculated using Kaplan-Meier estimates, as the cumulative probability of progression-free survival. Estimate of PFS3 will each be presented with 95% CIs. Median progression-free survival and plots of PFS rates over time will also be presented, based on the Kaplan-Meier estimates.

Proportion of patients alive and progression free after 6 months

The same analysis as for PFS3 will be conducted for the PFS rate at 6 months (PFS6).

4.2.6 Overall survival

Kaplan-Meier plots will be presented by treatment arm. Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow-up, and those who have withdrawn consent will be provided. The summaries will be produced for Korean and non-Korean.

In addition, duration of follow-up will be summarized using medians:

- In all patients: Time from first dosing to the date of death (i.e. overall survival) or to the date of censoring for censored patients regardless of treatment arm.

Proportion of patients alive at 6 months from randomization/enrolment

The OS6 will be defined as the Kaplan-Meier estimate of OS at 6 months.

Proportion of patients alive at 12 months from randomization/enrolment

The OS12 will be defined as the Kaplan-Meier estimate of OS at 12 months.

4.2.7 Change in tumor size

The absolute values, change in TL tumor size from baseline and percentage change in TL tumor size from baseline will be summarized using descriptive statistics and presented at each timepoint for each treatment arm. The best change in target lesion tumor size from baseline, (where best change in target lesion size is the maximum reduction from baseline or the minimum increase from baseline in the absence of a reduction) will also be summarized and presented for each treatment arm.

Tumor size will also be presented graphically using waterfall plots for each treatment arm, to present each subject's best percentage change in tumor size as a separate bar, with the bars ordered from the largest increase to the largest decrease. Reference lines at the +20% and -30% change in tumor size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. Additional waterfall plots showing percentage change in tumor size at specific time points may be produced if it is felt that these are warranted to provide greater clarity.

4.2.8 Healthcare resource use (Not applicable)

4.2.9 Safety data

Safety and tolerability data will be presented by treatment arm using the safety population. Safety data will be summarized only. No formal statistical analyses will be performed on the safety data.

Data from all cycles of treatment will be combined in the presentation of safety data. AEs (both in terms of MedDRA preferred terms and CTCAE grade) will be listed individually by patient. The number of patients experiencing each AE will be summarized by treatment arm and CTCAE grade. Additionally, data presentations of the rate of AEs per person-years at risk

may be produced. Any safety summaries examining retreatment with MEDI4736 monotherapy and MEDI4736 + tremelimumab combination therapy will be produced separately.

Other safety data will be assessed in terms of physical examination, clinical chemistry, hematology, vital signs, and ECGs. Exposure to MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy will be summarized. Time on study and MEDI4736 monotherapy, MEDI4736 + tremelimumab combination therapy dose delays/interruptions will also be summarized.

The following sections describe the planned safety summaries. However, additional safety tables may be required to aid interpretation of the safety data.

Adverse events

All AEs, both in terms of current MedDRA preferred term and CTCAE grade, will be summarized descriptively by count (n) and percentage (%) for each treatment group. The current MedDRA dictionary will be used for coding. The majority of the AE summaries, unless stated otherwise, will be based on TEAEs. Any AE occurring before study treatment (i.e. before the administration of the first dose on Study Day 1) will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'. However, any AE occurring before the administration of the first dose on Study Day 1 that increases in severity after the first dose will be regarded as treatment emergent and thus will be included in the majority of summary tables.

AEs observed up until 90 days following discontinuation of the study treatment or until the initiation of the first subsequent anti-cancer therapy following discontinuation of treatment (whichever occurs first) will be used for reporting of all of the AE summary tables. This will more accurately depict AEs attributable to study treatment only as a number of AEs up to 90 days following discontinuation of treatment are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, all of the AE summaries will also be produced containing AEs observed up until 90 days following discontinuation of the study treatment (i.e. without taking subsequent anti-cancer therapy into account).

A selection of AE summaries may also be produced containing AEs (by system organ class and preferred term) observed from the initiation of the first subsequent anti-cancer therapy following discontinuation of study treatment until 90 days following discontinuation of the study treatment (i.e. summarising those AEs experienced by patients taking subsequent therapy during the AE collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of AEs observed warrant the inclusion of such outputs for interpretational purposes. Any data post 90 days last dose for study treatment will be presented in a separate summary that presents any events that occur prior to dosing or starting more than 90 days after discontinuing treatment.

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved) and investigator's assessment of severity and relationship to study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from any on episode level summaries which may be produced).

Summary information (the number and percent of patients by system organ class and preferred term separated by treatment group) will be tabulated for:

- All AEs
- All AEs causally related to study medication (as determined by the reporting investigator)
- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related to study medication (as determined by the reporting investigator)
- AEs with outcome of death
- AEs with outcome of death causally related to study medication (as determined by the reporting investigator)
- AEs by outcome
- All SAEs
- All SAEs causally related to study medication (as determined by the reporting investigator)
- SAEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, causally related to study medication (as determined by the reporting investigator)
- AEs leading to hospitalization
- AEs leading to dose delay of study medication
- Other significant AEs
- Other significant AEs causally related to study medication (as determined by the reporting investigator)

- Immune mediated AEs (as determined by the reporting investigator)

- Infusion reaction AEs (as determined by the reporting investigator)

An overall summary of the number and percentage of patients will be presented In addition, a truncated AE table of most common AEs and another table showing most common AEs with CTCAE grade 3 or higher, showing all events that occur in at least 5% of patients overall will be summarized by preferred term, by decreasing frequency. This cut-off may be modified after review of the data. When applying a cut-off (i.e., x %), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency of 4.9% will not appear if a cut-off is 5%). Summary statistics showing the time to onset and the duration of the first AE may also be presented as appropriate.

Each AE event rate (per 100 patient years) will also be summarized by preferred term within each system organ class for the output summarizing all AEs. For each preferred term, the event rate is defined as the number of patients with that AE divided by the total drug exposure of patients at risk of AE. The denominator is calculated as the total over each patient of days from first dose to the earlier of the date of onset of the event or the last day of study medication.

Summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, system organ class, preferred term and treatment group.

Fluctuations observed in CTCAE grades during study will be listed for those AEs which are CTCAE ≥ 3.

In addition, all AEs will be listed.

Death

A summary of all deaths will be provided with number and percentage of patients by treatment group, categorized as:

1. Total number of death (including deaths within 90 days after last dose of study medication and deaths occur outside of 90 days post dosing window)

2. Death related to disease under investigation ONLY, as determined by investigator (including deaths within 90 days after last dose of study medication and deaths occur outside of 90 days post dosing window)

3. TEAE with outcome of death ONLY and onset date prior to initiation of subsequent anti-cancer therapy

4. AE with outcome of death ONLY and onset date falling after 90 days follow-up period or initiation of subsequent anti-cancer therapy (whichever is earlier) (1st summary)

5. Death related to disease under investigation, as determined by the investigator, and with TEAE with outcome of death and onset date prior to initiation of subsequent anti-cancer therapy

6. Death related to disease under investigation, as determined by the investigator, and with AE with outcome of death and onset date falling after 90 days follow-up period or initiation of subsequent anti-cancer therapy (whichever is earlier)

7. Death occurred 90 days after last dose or initiation of subsequent anti-cancer therapy (whichever is earlier), and unrelated to AE or disease under investigation

8. Patients with unknown reason for death.

9. Other deaths

Summaries 1, 2, 8 and 9 above will be produced twice: firstly accounting for subsequent therapy and, secondly, without taking subsequent therapy into consideration.

This summary will include AEs with start date up to 90 days after last dose or up to date of subsequent therapy, whichever occurs first.

Adverse events of special interest

Preferred terms used to identify adverse events of special interest (as defined in Section 3.3.1) will be listed before DBL and documented in the Study Master File. Grouped summary tables of certain MedDRA preferred terms will be produced and may also show the individual preferred terms which constitute each AESI grouping. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

Summaries of the above-mentioned grouped AE categories will include number (%) of patients who have:

- At least one AESI presented by outcome
- At least one AESI causally related to study medication
- At least one AESI leading to discontinuation of study medication

A summary of total duration (days) of AESI will be provided for events which have an end date and this will be supported by summaries of ongoing AESIs at death and, separately, at data cut-off.

Additionally, there will be several summaries of AESIs requiring concomitant treatment, and particularly the relationship of AESIs to the use of immunosuppressive agents (i.e., depicting

which AESI triggered immunosuppressive use) and, separately, to the use of immunosuppressive agents at high doses.

Laboratory assessments

Data obtained up until the 3 months (+1 week for a late assessment, i.e. 3*30+7=97 days) following discontinuation of treatment or until the initiation of the first subsequent therapy will be used for reporting. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 97 days following discontinuation of study treatment are likely to be attributable to the subsequent therapy.

However, to assess the longer term toxicity profile, summaries of laboratory data will also be produced containing data collected up until 97 days following discontinuation of study treatment irrespective of whether subsequent therapy was administered during that period.

A small selection of summaries of laboratory data may also be produced containing data from initiation of the first subsequent therapy following discontinuation of study treatment until 97 days following discontinuation of study treatment (i.e. summarizing the laboratory data collected on patients taking subsequent therapy during the safety collection follow-up window post discontinuation of study treatment). These outputs will only be produced if the number of laboratory toxicities observed warrant the inclusion of such outputs for interpretational purposes. Any data post 97 days last dose for study treatment will not be summarized.

Data summaries will be provided in International System (SI) of units.

Scatter plots (shift plots) of baseline to maximum value/minimum value (as appropriate) on treatment (i.e. on-treatment is defined as data collected between the start of treatment and the relevant follow-up period following the last dose of study treatment) may be produced for certain parameters if warranted after data review (e.g. alanine aminotransferase [ALT], aspartate aminotransferase [AST], alkaline phosphatase [ALP], total bilirubin, corrected calcium, lactate dehydrogenase [LDH], magnesium, sodium, potassium, glucose, creatinine, urea nitrogen, and thyroid stimulating hormone [TSH], free triiodothyronine [fT3] and free thyroxine [fT4]).

Scatter plots (shift plots) of baseline to minimum value on treatment (defined as above) may be produced for certain parameters if warranted after data review (e.g. hemoglobin, lymphocyte (count, absolute), neutrophils (count, absolute), platelet count, albumin, total protein, corrected calcium, magnesium, sodium, potassium, glucose and TSH, fT3 and fT4).

Box-plots of absolute values by week, and box-plots of change from baseline by week, may be presented for certain parameters if warranted after data review (e.g. hemoglobin, neutrophil (count, absolute), lymphocyte (count, absolute), platelet (count), AST, ALT, ALP, total bilirubin, albumin, total protein, corrected calcium, LDH, sodium, potassium, creatinine, urea nitrogen and TSH, fT3 and fT4).

For continuous laboratory assessments including the thyroid test parameters TSH, fT3 and fT4, absolute value, change from baseline and percentage change from baseline will be summarized using descriptive statistics at each scheduled assessment time by actual treatment group.

Shift tables for laboratory values by worst CTC grade will be produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change will be produced. The laboratory parameters for which CTC grade shift outputs will be produced are:

- Hematology: hemoglobin, leukocytes, lymphocytes (absolute count), neutrophils (absolute count), platelets

- Clinical chemistry: ALT, AST, ALP, total bilirubin, albumin, magnesium – hypo and – hyper, sodium – hypo and – hyper, potassium – hypo and – hyper, corrected calcium – hypo and – hyper, glucose – hypo and – hyper, creatinine

For the parameters with no CTCAE grading that are listed in the CSP, shift tables from baseline to worst value on-treatment will be provided.

Additional summaries will include a shift table for urinalysis (bilirubin, blood, glucose, ketones, protein) comparing baseline value to maximum on-treatment value.

Shift tables showing baseline to maximum and baseline to minimum will be produced for TSH, fT3 and fT4.

Liver Enzyme Elevations and Hy's law

The following summaries will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study

  - ALT ≥ 3x – ≤ 5x, > 5x – ≤ 8x, > 8x – ≤10x, > 10x – ≤ 20x, and > 20x Upper Limit of Normal (ULN) during the study

  - AST ≥ 3x – ≤ 5x, > 5x – ≤ 8x, > 8x – ≤10x, > 10x – ≤ 20x, and > 20x ULN during the study

  - Total bilirubin ≥ 2x – ≤ 3x, > 3x – ≤ 5x, > 5x ULN during the study

  - ALT or AST ≥3x – ≤ 5x, >5x – ≤ 8x, > 8x – ≤10x, > 10x – ≤ 20x, and > 20x ULN during the study

  - ALT or AST ≥3x ULN and total bilirubin ≥2x ULN during the study (Potential Hy's law): The onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation

- Narratives will be provided in the CSR for patients who have ALT ≥ 3x ULN plus total bilirubin ≥ 2x ULN or AST ≥ 3x ULN plus total bilirubin ≥ 2x ULN at any visit.

For Hy's Law criteria to be met, the elevation in transaminases must precede or be coincident with the elevation in total bilirubin (i.e. on the same day), but there is no specified timeframe within which the elevations in transaminases and total bilirubin must occur.

Liver biochemistry test results over time for patients with elevated ALT or AST (i.e. ≥ 3x ULN), and elevated total bilirubin (i.e. ≥ 2x ULN) (at any time) will be plotted. Individual patient data where ALT or AST (i.e. ≥ 3x ULN) plus total bilirubin (i.e. ≥ 2x ULN) are elevated at any time will be listed also.

Plots of ALT and AST vs. total bilirubin by treatment group will also be produced with reference lines at 3xULN for ALT, AST, and 2xULN for total bilirubin. In each plot, total bilirubin will be in the vertical axis.

ECGs

Overall evaluation of ECG is collected at screening and as clinically indicated while on study treatment in terms of normal or abnormal, and the relevance of the abnormality is termed as "clinically significant" or "not clinically significant". A shift table of baseline evaluation to worst evaluation will be produced.

Vital signs

Vital signs data obtained up until the 30-day safety follow-up visit will be included in the summary tables.

Box plots for absolute values and change from baseline by week, may be presented for certain vital signs parameters if warranted after data review.

Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate, temperature, respiratory rate and weight) will be summarized over time in terms of absolute values and changes from baseline at each scheduled measurement by treatment group.

Time to subsequent therapy from discontinuation of study treatment

A descriptive summary will be produced for time to subsequent therapy from discontinuation of study treatment. This summary is supportive of the Adverse Event and Laboratory data outputs.

Physical examination

All individual physical examination data will be listed only.

Other safety data

Data from positive pregnancy tests will be listed only.

4.2.10 WHO performance status

All WHO/ECOG performance status will be summarized over time.

4.2.11 PK/PDx data

Pharmacokinetic concentration data will be listed for each patient and each dosing day, and a summary provided for all evaluable patients in the PK analysis population.

PK modeling will be performed outside of the CSR.

Immunogenicity analysis

Immunogenicity results will be listed by patient and a summary will be provided of the number and percentage of patients who develop detectable anti-MEDI4736 and anti-tremelimumab antibodies based on the safety population. The immunogenicity titer and neutralizing ADA data will be listed for samples confirmed positive for the presence of anti-MEDI4736 antibodies and/or anti-tremelimumab antibodies.

The effect of immunogenicity on PK, PDx, efficacy and safety will be evaluated if data allow and such analysis will be performed outside of the CSR.

4.2.12 PK/PDx relationship

If the data are suitable, the relationship between PK exposure and efficacy/safety parameters may be investigated graphically or using an appropriate data modeling approach. This will be performed outside of the CSR

4.2.13 Biomarker data

The relationship of PD-L1 expression and if applicable, of exploratory biomarkers to ORR, DoR, DCR, PFS, PFS3 and PFS6 will be presented for a subset of patients in the ITT population who are evaluable for each biomarker.

This will be assessed using similar summary and graphical representations to those that are outlined for the efficacy outputs in Section 4.2.2 to Section 4.2.5.

PD-L1 expression determined by IHC assay will be reported in the CSR. Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the CSR in a separate report.

4.2.14 Demographic and baseline characteristics data

The following will be summarized for all patients in the FAS (unless otherwise specified) by treatment group:

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations

- Inclusion in analysis populations
- Demographics (derived age, derived age group[<50, ≥ 50 - < 65, ≥ 65 - < 75 years and ≥ 75 years], sex, race and ethnicityPatient characteristics at baseline (height, weight, weight group[<70kg, 70kg to 90 kg, >90kg],
- Patient recruitment by country and center
- Previous disease-related treatment modalities
- Number of regimens of previous chemotherapy at baseline
- Disease characteristics at baseline (World Health Organization performance status, primary tumor location, histology type, tumor grade, American Joint Committee on Cancer staging, pancreas cancer location and overall disease classification, best response to previous therapy , time from informed consent to first dose[≤14 days, 14 days < to ≤21 days, 21 days < to ≤42 days, >42 days]). It will be provided for both overall and by Korea status (i.e. Korea and Non Korea).
- Extent of disease at baseline
- Disease related medical history (past and current combined)
- Relevant surgical history
- Physical examination at baseline
- Time from most recent disease progression to start of study treatment
- Disallowed concomitant medications
- Allowed concomitant medications
-
- Nicotine use, categorized (never, current, former) and descriptive statistics by type of substance use
- Stratification factors as per IVRS/IWRS and eCRF data

The AZ drug dictionary will be used for concomitant medication coding.

Patient disposition data will also be summarized at the time of OS analysis.

4.2.15 Treatment exposure

The following summaries related to study treatment will be produced for the Safety Analysis Set by randomized treatment group, for the initial treatment period:

- Total exposure (weeks) (*) .
- Actual exposure (weeks) (*) .
- Total number of cycles received (*).
- Reasons for dose delays and infusion interruptions of MEDI4736 and tremelimumab. Dose interruptions will be based on investigator initiated dosing decisions.
- Number of infusions received (*).
- RDI (relative dose intensity) of MEDI4736 and tremelimumab.

In the combination therapy arm, all descriptions above will be provided separately for each molecule and, for those having (*), in total, where the maximum in either molecule will be reported. The total and actual exposures will also be expressed as total treatment years, defined as the total across all patients in the analysis set.

Data collected during the re-treatment will only be listed.

Time on study

Time on study is defined as the [study discontinuation date – randomization date +1] and will be summarized by treatment arm.

4.2.16 Subsequent therapy

Subsequent therapies received after discontinuation of study treatment will have summaries produced by treatment group, together with number of regimens received.

5. TIMING OF ANALYSES

5.1 Part A: Interim Analysis

An interim analysis when 15 evaluable patients in each arm are observed in Part A will use predictive probability to determine whether either arm may stop (based on <10% predictive probability of meeting the minimum criteria to initiate Part B). Should it become obvious that the continuation criteria have been met prior to completion of enrolment of Part A, then the decision to initiate Part B (either Expansion or the RCT) may be made early.

In each arm, once 15 patients have completed 2 post-baseline tumor assessments (at Week 6 and 12), predictive probabilities will be calculated to assess the chance of observing at least 5 out of 30 responses (>15% ORR), and (separately) the chance of observing at least 15/30 DCR at 12 weeks (≥15 DCR12) in that arm. Only data from the 15 evaluable patients will initially support the predictive probabilities.

However in the event of treatment discontinuation, the impact of such patients on these predictive probabilities will be considered.

This will make use of a Beta-Binomial posterior predictive distribution, where at the outset a non-informative prior is assumed for the ORR or DCR12 (Beta[1,1]), and a Beta-Binomial distribution can then be used to calculate the probability of achieving the required ORR or DCR from 30 evaluable patients given data observed.

If predictive probability is below 10% for both endpoints, enrolment into the arm in which this happens will be stopped. A different decision outside of these criteria may be made, if the totality of the data supports it. Recruitment will continue while predictive probability is evaluated. If enrolment into an arm is to be stopped based on <10% predictive probability, no new patients will be recruited, but patients who are already on study will continue in accordance with study guidelines. If enrolment is fast enough so that, by the time 15 patients are evaluable, all 30 patients per arm have already been randomized, then the predictive probability calculations will not be carried out.

5.2 Primary analyses

For both Part A and Part B: Expansion, a primary analysis will be performed when all subjects have completed the Week 18 visit or discontinued treatment. For Part A analysis, all data present at that time will be used for decision making on format for Part B.

6. CHANGES OF ANALYSIS FROM PROTOCOL

6.1 Disease control rate

As per CSP Section 8.4.1.3:

DCR at 3 or 12 months is defined as the percentage of patients who have a best objective response (BoR) of CR or PR in the first 3 or 12 months, respectively, or who have demonstrated SD for a minimum interval of 12 or 52 weeks, respectively (-7 days, ie, 105 or 357 days, respectively), following the start of study treatment.

For DRC at 3 months, the “105 days” mentioned for minimum duration for SD will be replaced with “77 days” (see Section 3.2.2.8) since a 13-weeks minimum duration for SD (minus 1 to allow for an early assessment) corresponds to (13-1)*7=84 days.

In addition, DCR at 6 months will be provided to get an intermediate estimate.

6.2 Protocol amendment

As per CSP Section 1.4:

The format of Part B will be determined based on the responses seen in the first 30 patients in each arm in Part A. A protocol amendment will be made if criteria are met to proceed to Part B.

In order to clarify the sentence above, a protocol amendment is needed only if the data suggest that Part B: RCT is to be initiated.

6.3 Part B: Expansion design

As per CSP Synopsis,

This study will consist of Part A, Lead-in, as well as a possible Part B, non-randomized expansion (Part B: Expansion) or a possible Part B, randomized controlled study (Part B: RCT).

In order to clarify the sentence above, Part B: Expansion will be randomized if both arms initiated in Part A go to Part B: Expansion, but if only one goes, Part B: Expansion will not be randomized.

6.4 ePRO data

ePRO data will be collected only if Part B: RCT is initiated.

7. REFERENCES

Aaronson et al 1993
Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85(5):365-76.

Fitzsimmons and Johnson 1998
Fitzsimmons D, Johnson CD. Quality of life after treatment of pancreatic cancer. Langenbecks Arch Surg 1998;383(2):145-51.

Fitzsimmons et al 2005
Fitzsimmons D, Kahl S, Butturini G, van Wyk M, Bornman P, Bassi C, et al. Symptoms and quality of life in chronic pancreatitis assessed by structured interview and the EORTC QLQ-C30 and QLQ-PAN26. Am J Gastroenterol 2005;100(4):918-26.

Rahma et al 2013
Rahma OE, Duffy A, Liewehr DJ, Steinberg SM, Greten TF. Second-line treatment in advanced pancreatic cancer: a comprehensive analysis of published clinical trials. Ann Oncol 2013;24:1972-9

8. APPENDIX (NOT APPLICABLE)