AstraZeneca

Statistical Analysis Plan

Study Code D4190C00022

Edition Number 8.0

Date 27 November 2020

A Study of Safety, Tolerability, and Clinical Activity of Durvalumab and Tremelimumab Administered as Monotherapy, or Durvalumab in Combination with Tremelimumab or Bevacizumab in Subjects with Advanced Hepatocellular Carcinoma

Version History

Version	Statistician	Date	Notes / Revision
1.0	PPD	8-Dec-2015	Original version
2.0		20-Dec-2016	Updated per protocol amendment 1 and 2.
4.0		24-May-2018	Updated per protocol amendment 3: • Updated Study Flow Diagram (Figure 3.1.1-1) to reflect additional treatment arms. • Added Part 2B Arm D fixed-dose combo therapy • Added Part 3 fixed-dose combo therapy (Arms A, B, C, and D), MEDI4736 and tremelimumab mono therapy. • Updated sample size/power calculation for Part 3. • Removed the reference to RP2D • Removed appendix of tumor response assessment using irRECIST criteria. • Updated Section 4 Interim Analysis by adding second interim analysis. • Added protocol deviations (Section 3.2.2) • Added Deviation of RECIST visit responses (Section 3.4.2). • Added time to response analysis (Section 3.4.1.4) Updated to protocol amendment 4: • Based on Protocol Amendment 4 approved on 19-Dec-2018. • FAS analysis set was defined for the efficacy analyses and summarizing of demography for both Part 2 and Part 3. • Updated Section 3.7.5 clinical laboratory data analysis following HIMALAYA study. • Child-Pugh scores analysis method was added. • Added definitions for PD-L1 expression (+, -, NE) at randomization and for subgroup analysis (low vs high).
5.0		25-Feb-2019	• Changes related to protocol amendments 5 and 6. • Additional text about interim analyses and multiple testing. • Removal of an unnecessary sentence in BICR section.
6.0		16-Oct-2019	• Reorganized sections slightly to match better standard AZ SAP template for late phase studies. • Section 2.1: renamed "As-treated population" to "Safety analysis set", named “Full Analysis Set” as the primary population for all efficacy analyses. Only ORR will be analyzed using the Response Evaluable Population. • Updated list of abbreviations. • Section 3.1.1: Updated RECIST baseline to be relative to randomization for Parts 2A and 3, and relative to first dose

			of study drug in Parts 1, 2B and 4. Included more details and clarifications about baseline calculations. • Section 3.3.1.3: Added details to the DCR definition. • Section 3.3.1.8: Updated derivation for censoring date for OS and added notes about method of calculation in case of absence of survival calls. • Section 3.3.1.5: Adjusted PFS censoring rules to more comply with a global standard, updated definition to be relative to randomization for Parts 2A and 3, and relative to first dose of study drug in Parts 1, 2B and 4, added explanations about two missing visits calculation. • Section 3.3.1.1: Updated definition to be relative to randomization for Parts 2A and 3, and relative to first dose of study drug in Parts 1, 2B and 4. CCI [redacted] • Section 1.2: added a paragraph about study parts, and paragraphs explaining retreatment. • Section 1.3: explained 6 study parts. CCI [redacted] • Section 2.2: Updated Important Protocol Deviations list and added paragraphs explaining analysis of deviations. • Section 3.1: Added limited number of listing to produce and added details about naming of treatment arms in TFLs, display of decimal places, • Section 3.2.2: Updated analyses to be done within demographics and baseline characteristics. • Section 3.1.2: Clarified window for baseline. • Section 3.2.3: Added analyses to be done for concomitant medications. • Sections 3.2.4.1, 3.2.4.2: Added clarifications about methods of calculations and analyses to be done. • Added notes in the document that both confirmed and unconfirmed ORR will be calculated. • Section 3.3.2: Updated definition of baseline for RECIST and made cosmetic corrections. • In all document replaced references to Response evaluable population with Full analysis set, and references to As-treated population with Safety analysis set. • Section 3.3.3.10: Updated PDL1 status section in line with latest agreements and similarly to HIMALAYA study.

			• CCI [redacted] • Section 3.6.1: Updated to mention that DLT will be summarized for Part 1A patients only. • Section 4: Added further clarifications about interim analyses. • Section 5: Added a new section to describe changes to analysis from protocol.
7.0	PPD	16-Mar-2020	• Section 1.2: Updated durvalumab abbreviation, explanation of dosing. • Section 2.1: Added explanation of analysis sets during interim analyses. • Section 3.1: Revised display order of treatment arms. • Section 3.2.2: Updated contents of some Demographics and Baseline Characteristics summaries. • Section 3.2.3: Added definitions and date imputation rules for prior and concomitant medications, updated drug dictionary • Section 3.2.4.1: Updated tremelimumab monotherapy exposure calculations; redefined dose delays as cycle delays • Section 3.3.1: Added landmark OS efficacy endpoints • Section 3.3.1.2: Added more information to define objective response rate • Section 3.3.4: Added subgroup analyses • Section 3.6: Defined “on treatment” or treatment-emergent for all safety analyses • Section 3.6.6.2: Defined QT interval corrected for using Fridericia’s formula • Section 3.6.6.4: Dropped change from baseline Child-Pugh score and class summaries • Section 4: Added explanation of Interim Analysis 6
8.0		27-Nov-2020	• Updated abbreviations list • Section 1.2: amended wording to imply that in Part 4 durva is fixed dose, and beva is weight-based; removed references to IA for Part 4; removed references to an external SAP for the China tail • Section 2.1: clarified DLT evaluable set definition

			• Section 3.1: defined display of treatment arms in tables and figures; updated list of endpoints in efficacy response data listing; added a paragraph about safety data handling; added a definition of China cohort; added date imputation rules for AEs and ConMeds. • Section 3.1.2: amended visit windows labels; removed paragraph about only showing visits with meaningful data, because this had not been implemented for Study 22, added windows for assessments following non-standard schedule. • Section 3.1.3: added a note that China tail data will be reported in global study CSR, not separately, and that DCO will be same like global. • Section 3.2.2: added a reference to the new section about viral status derivation; amended baseline tables list • Section 3.2.3: amended section to match TA SAP and HIMALAYA approach. • Section 3.2.4.1: amended Total exposure definition in line with comments; removed categorization which is not used in tables. • Section 3.2.4.2: corrected intended cumulative dose definition, removed categories of RDI to summarize; added a note that retreatment period refers only to Parts 1-3; corrected definition of intended number of cycles to Treme mono arm; clarified intended dose calculation for weight based dosing with not completed visits; clarified that for Part 4 there will be no split by initial and retreatment phase. • Section 3.3.1: removed DCR-24w, TTP-6m and PFS-6m, and change from baseline in tumor size from endpoints. • Section 3.3.1.1: amended BoR definition • Section 3.3.1.2: added a note that unconfirmed ORR will be additionally reported in CSR • Section 3.3.1.3: deleted DCR-24w • Section 3.3.1.5: deleted PFS-6m, clarified censoring rules for PFS • Section 3.3.1.6: added a note about subset • Section 3.3.1.7: removed TTP-6m • Section 3.3.1.8: removed subsequent anticancer treatment page from list to check for last known date alive; added a note about calculation of OS-18m and OS-24m. Clarified imputed death date rule.

			• Section 3.3.2.6: added a note about availability of BICR data • Section 3.3.2.7: removed DCR-24w, TTP-6m and PFS-6m from list of endpoints. • Section 3.3.3: moved part about methods of analyses to later sections; amended list of summaries to perform. • Sections 3.3.3.1-3.3.7: added a list of endpoints to derive • Section 3.3.3.3: deleted DCR-24w • Section 3.3.3.5: removed PFS-6m • Section 3.3.3.6: amended analysis to be only descriptive • new Section 3.3.3.11: viral status definition • Section 3.3.4: amended subgroups list • CCI [redacted] • Section 3.6: clarified that date of first subsequent anticancer systemic therapy should be checked. • Section 3.6.1: DLT evaluation will be displayed overall for all patients, not by dose cohort. • Section 3.6.2: added an explanation about AEs requiring steroids and missing dates imputation. • Section 3.6.3: added definitions of AEPI and imAE. • new Section 3.6.4 added to reflect what is summarized in standard tables. • new Section 3.6.5 added to reflect what is summarized in standard tables. • Section 3.6.7: updated section about liver tests elevations and thyroid tests summaries to reflect what is summarized in standard tables. • Section 3.6.8.2 clarified visit assessment calculation for multiple assessments per visit/ day. • Section 4: removed references to IA for Part 4 • Section 5: updated list of changes to analysis from protocol

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

Abbreviation or Specialized Term	Definition
CCI	
AE	adverse event
AEPI	adverse event of possible interest
AESI	adverse event of special interest
CCI	
AHQ	ad-hoc queries
AJCC	American Joint Committee on Cancer
ALP	alkaline phosphatase
ALT	alanine transaminase
AST	aspartate transaminase
ATC	anatomical therapeutic chemical
AZ	AstraZeneca
AZDD	AstraZeneca drug dictionary
BCLC	Barcelona-Clinic Liver Cancer
BICR	Blinded Independent Central Review
BMI	body mass index
BoR	best objective response
CI	confidence interval
CR	complete response
CRF	case report form
CSP	clinical study protocol
CT	computed tomography
CTCAE	common terminology criteria for adverse events
DBL	database lock
DCO	data cut-off
DCR	disease control rate
DCR-24W	disease control rate at 24 weeks
DLT	dose-limiting toxicity
DoR	duration of response
ECG	electrocardiography
ECOG	Eastern Cooperative Oncology Group
EHS	extrahepatic spread
FAS	full analysis set
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
IA	interim analysis
IC	immune cells
ICF	informed consent form
IP	investigational product
IV	intravenous
IWRS	interactive wen response system

Abbreviation or Specialized Term	Definition
MedDRA	medical Dictionary for Regulatory Activities
CCI	[redacted]
MRI	magnetic resonance imaging
MVI	Macrovascular invasion
NA	not applicable
nAB	neutralizing antibody
NCA	non-compartmental analysis
NCI	National Cancer Institute
NE	not evaluable
NED	no evidence of disease
ORR	objective response rate
OS	overall survival
OS-1y	overall survival at 1 year
PD	progressive disease
PD	pharmacodynamics
PD-L1	programmed cell death ligand 1
PFS	progression-free survival
PFS-6m	progression-free survival at 6 months
CCI	[redacted]
PR	partial response
PT	Preferred term
Q3W	every three weeks (Q = *quoque*, Latin)
Q4W	every four weeks (Q = *quoque*, Latin)
Q8W	every eight weeks (Q = *quoque*, Latin)
Q9W	every nine weeks (Q = *quoque*, Latin)
Q12W	every twelve weeks (Q = *quoque*, Latin)
QTcB	QT corrected using Bazett's formula
QTcF	QT corrected using Fridericia's formula
RDI	relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	serious adverse event
SAP	statistical analysis plan
SAS	statistical Analysis System (Cary, NC, USA)
SD	stable disease
SMQ	standardised medDRA query
SOC	system organ class
TC	tumor cells
TEAE	treatment-emergent adverse event
TFL	tables, figures and listings
TL	target lesion
TSH	thyroid stimulating hormone
TTP	time to progression

Abbreviation or Specialized Term	Definition
TTP-6m	time to progression at 6 months
TTR	time to response
ULN	upper limit of normal
VEGFR TKI	vascular endothelial growth factor receptor tyrosine kinase inhibitor

1 STUDY DETAILS

This document describes the statistical methodology and summaries in accordance with Clinical Study Protocol Amendment 6 (4-Jan-2019) for Study D4190C00022, an investigation of durvalumab and tremelimumab administered as monotherapy, or durvalumab in combination with tremelimumab or bevacizumab in subjects with advanced hepatocellular carcinoma (HCC). The main portion of this document details the statistical summaries relating to each study objective and describes the general conventions and definitions that will be used. In addition, a separate document consisting of a set of templates and specifications for tables, figures and listings (TFL) will be created to complement this document.

1.1 Study Objectives

1.1.1 Primary Study Objective

1. To assess the safety and tolerability of durvalumab and tremelimumab administered as monotherapy, and durvalumab in combination with tremelimumab or bevacizumab in subjects with advanced HCC.

1.1.2 Secondary Study Objectives

1. To evaluate the efficacy of durvalumab and tremelimumab administered as monotherapy, and durvalumab in combination with tremelimumab or bevacizumab in subjects with advanced HCC.
2. To evaluate the relationship between programmed cell death ligand 1 (PD-L1) expression in the tumor microenvironment and measures of clinical outcome of durvalumab and tremelimumab administered as monotherapy, and durvalumab in combination with tremelimumab or bevacizumab in subjects with advanced HCC.

1.1.3 Exploratory Study Objectives

CCI

CCI

1.2 Study Design

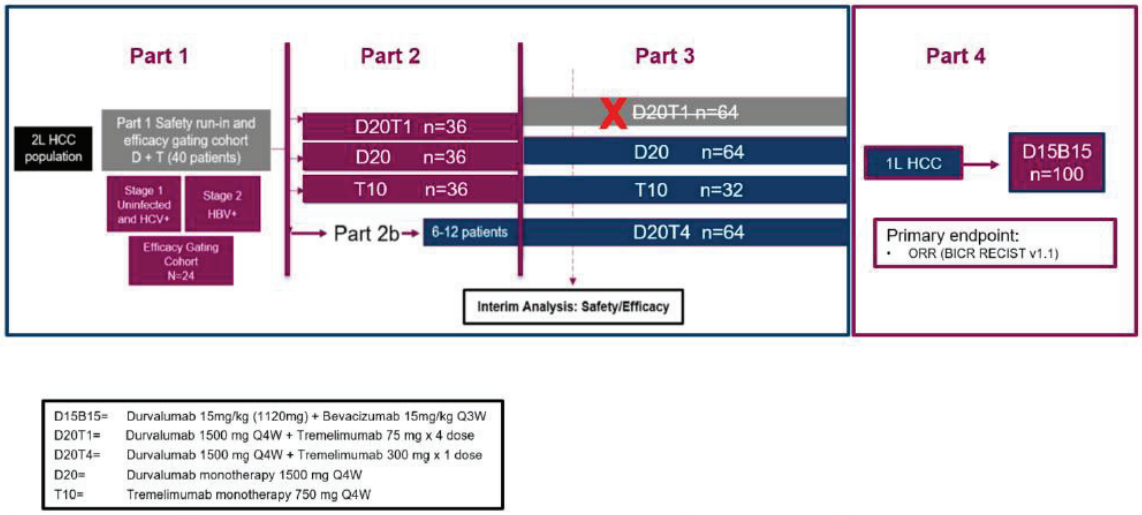
This is a multicenter, open-label, stratified, randomized study designed to evaluate the safety, tolerability, and clinical activity of durvalumab and tremelimumab administered as monotherapy, and durvalumab in combination with tremelimumab or bevacizumab, in subjects with advanced HCC. Approximately 655 subjects will be screened to enroll approximately 456 subjects globally, including approximately 12 subjects in Part 1A, approximately 24 subjects in Part 1B, approximately 108 subjects in Part 2A, approximately 12 subjects in Part 2B, approximately 200 subjects in Part 3, and approximately 100 subjects in Part 4.

Subjects will be treated in Part 1A, Part 1B, Part 2A, Part 2B, Part 3 and Part 4 of the study. Subjects in Part 1A, Part 1B, and Part 2A will receive a weight-based dosing regimen of durvalumab and tremelimumab either as monotherapy or combination therapy. Subjects in Part 2B and Part 3 will receive a fixed dosing regimen of durvalumab and tremelimumab either as monotherapy or combination therapy. Subjects in Part 4 will receive a fixed dosing regimen of durvalumab and a weight-based dosing regimen of bevacizumab as a combination therapy. Details are described in Protocol Section 3.1.2 as well as Protocol Table 3.1.2-1.

The study will include several interim analyses and one final analysis. Details on interim analyses are provided in Section 4. Additional 'ad hoc' analyses beyond the scope of the SAP may be performed to support the design and development of other future HCC studies for internal purposes. Such analyses will be fully documented outside of the SAP. The study will be comprised of 6 parts (1A, 1B, 2A, 2B, 3 and 4) described below and illustrated in Figure 1.2-1.

Following protocol amendment 5, enrollment of Part 3 Arm A was closed. Patients randomized to Arm A before this protocol amendment, can continue on assigned study treatment (provided the investigator and patient think it is in the best interests of the patient) until confirmed progressive disease (PD) or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. If a patient has not completed or started all 4 doses of tremelimumab, the patient may either continue to complete the full schedule or continue with durvalumab monotherapy only.

Figure 1.2-1 Study Flow Diagram



D15B15=	Durvalumab 15mg/kg (1120mg) + Bevacizumab 15mg/kg Q3W
D20T1=	Durvalumab 1500 mg Q4W + Tremelimumab 75 mg x 4 dose
D20T4=	Durvalumab 1500 mg Q4W + Tremelimumab 300 mg x 1 dose
D20=	Durvalumab monotherapy 1500 mg Q4W
T10=	Tremelimumab monotherapy 750 mg Q4W

2L = second line; 1L = first line; HBV+ = hepatitis B virus positive; HCC = hepatocellular carcinoma; HCV+ = hepatitis C virus positive; D = Durvalumab; T = tremelimumab; B = Bevacizumab

Part 1A: Safety Run-in with Durvalumab and Tremelimumab Combination Therapy

Immunotherapy-naïve subjects with advanced HCC who have progressed on, are intolerant of, or refused sorafenib-based therapy will be enrolled in Part 1A based on a risk-based staggered approach.

- **Stage 1:** Approximately 6 subjects with advanced uninfected or HCV+ HCC will be enrolled. Subjects will be administered the durvalumab 20 mg/kg and tremelimumab 1 mg/kg combination therapy every 4 weeks (Q4W) for 4 doses followed by durvalumab 20 mg/kg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. If the frequency of adverse events (AEs) meeting dose-limiting toxicity (DLT) criteria is ≥ 33% for a given viral status/type, then lower dose cohorts in that specific subpopulation may be explored depending on the type and severity of the toxicities seen at this combination dose.

- **Stage 2:**
  - *HBV+ Cohort:* Enrollment of approximately 6 additional subjects with advanced HBV+ HCC may start after the first 3 subjects in Stage 1 have been observed on study for at least 4 weeks. Subjects will be administered the durvalumab 20 mg/kg and tremelimumab 1 mg/kg combination therapy Q4W for 4 doses followed by durvalumab 20 mg/kg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. If the frequency of AEs meeting DLT criteria is ≥ 33%, then Dose Level -1 (i.e., durvalumab 15 mg/kg and tremelimumab 1 mg/kg combination therapy Q4W followed by durvalumab 15 mg/kg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met) may be explored depending on the type and severity of the toxicities seen at this combination dose.
  - *Dose for HBV+ HCC Subjects Administered the Durvalumab and Tremelimumab Combination:* The durvalumab 20 mg/kg and tremelimumab 1 mg/kg combination therapy Q4W followed by durvalumab 20 mg/kg monotherapy Q4W will be the dose evaluated for HBV+ subjects in Part 1B and Part 2A if the following criteria are met: (1) all 6 subjects have been observed for at least 4 weeks and the DLT frequency is < 33% (i.e., < 2 of 6 subjects); AND (2) at least 3 of the 6 subjects have been observed for 6 weeks. If the DLT frequency for the Stage 2 cohort is ≥ 33% (i.e., ≥ 2 of 6 subjects), then Dose Level -1(i.e., durvalumab 15 mg/kg and tremelimumab 1 mg/kg combination therapy Q4W followed by durvalumab 15 mg/kg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met) will be the dose evaluated for HBV+ subjects in Part 1B and Part 2A assuming the DLT frequency at Dose Level -1 is < 33%.

Part 1B: Efficacy Gating Cohort for Durvalumab and Tremelimumab Combination Therapy

Immunotherapy-naïve subjects with advanced HCC who have progressed on, are intolerant of, or refused sorafenib-based therapy will be enrolled in Part 1B.

Approximately 24 subjects (uninfected, HBV infected, or HCV infected) will be enrolled in an efficacy gating cohort to determine if there is sufficient evidence of clinical activity to warrant opening enrollment to Part 2. An interim analysis will be conducted after Part 1A and Part 1B are fully enrolled and subjects have been followed for at least 16 weeks. Refer to Protocol Section 4.8.7 for details regarding the definition of sufficient evidence.

Subjects will be administered durvalumab 20 mg/kg and tremelimumab 1 mg/kg combination therapy Q4W followed by durvalumab 20 mg/kg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. After approximately 12 of the 24 subjects have been enrolled, enrollment of subjects who have

progressed on or are intolerant to sorafenib may be paused in favor of subjects who have refused sorafenib to ensure a minimum enrollment of 12 subjects who have refused sorafenib in Part 1 (Parts 1A and 1B). The total number of subjects with a specific viral status may be restricted in this cohort to ensure that all types of patients have an opportunity to enroll. For example, as enrollment proceeds if emerging data indicate that the majority of patients enrolling are all HBV infected, enrollment of this specific viral type of HCC may be paused to allow other types of subjects (uninfected and HCV infected) to enroll.

If during Part 1B of the study, ≥ 33% of subjects with a specific viral status type discontinue therapy for treatment-related toxicity, enrollment for that specific viral status type may be paused and study data will be reviewed to determine whether additional monitoring, alternate dose levels, or treatment schedules should be evaluated prior to further enrollment of subjects with that specific viral status type.

Part 2A: Randomized Arms Evaluating Durvalumab and Tremelimumab Combination Therapy, Durvalumab Monotherapy, and Tremelimumab Monotherapy

Immunotherapy-naïve subjects with advanced HCC who have progressed on, are intolerant of, or refused sorafenib-based therapy, will be stratified based on viral status (uninfected, HCV, or HBV) and PD-L1 expression (positive, negative, or non-evaluable).

PD-L1 expression was defined as: (1) positive if there were any tumor cells (TC) or immune cells (IC) staining, (2) negative if there was 0% TC and 0% IC staining, or (3) non-evaluable if the PD-L1 test failed, or the PD-L1 result is not available for a patient.

- Subjects may not enroll in any arm of Part 2A until sufficient evidence, as specified in Protocol Section 4.8.7 (in Part 1A and Part 1B), is observed.
- Subjects will be randomized 1:1:1 within each stratum to 1 of the 3 treatment arms, i.e., approximately 36 subjects (approximately 12 subjects/viral status type) per treatment arm:
  - Arm A: durvalumab and tremelimumab combination [D 20 mg/kg (or 15 mg/kg for Dose Level -1, if needed) + T 1 mg/kg Q4W ×4 doses followed by D 20 mg/kg Q4W]
  - Arm B: durvalumab monotherapy (D 20 mg/kg Q4W)
  - Arm C: tremelimumab monotherapy (T 10 mg/kg Q4W ×7 doses, then Q12W)

Subjects in Part 1A, Part 1B, and Arm A of Part 2A will receive durvalumab 20 mg/kg (or 15 mg/kg for Dose Level -1, if needed) in combination with tremelimumab 1 mg/kg Q4W for 4 doses. After completion of the initial 4 doses of combination therapy, subjects will receive durvalumab monotherapy at 20 mg/kg Q4W until any of the criteria outlined in Protocol Section

4.1.6 are met. The first durvalumab monotherapy dose at 20 mg/kg Q4W will be given 4 weeks after the final dose of durvalumab and tremelimumab combination therapy. Subjects who progress on durvalumab monotherapy may be retreated with the durvalumab and tremelimumab combination as described in Protocol Section 3.1.2.1.

Subjects in Arm B of Part 2A will receive durvalumab monotherapy at 20 mg/kg Q4W until any of the criteria outlined in Protocol Section 4.1.6 are met.

Subjects in Arm C of Part 2A will receive tremelimumab at 10 mg/kg Q4W for 7 doses and then Q12W until any of the criteria in Protocol Section 4.1.6 are met.

All subjects in Part 1A, Part 1B, and Part 2A will be evaluated for efficacy, and their disease status primarily analyzed according to RECIST v1.1. All subjects will be followed for survival until the end of study as defined in Protocol Section 6.3. CCI [redacted] Collection of noncancerous liver tissue at screening should also be attempted if it can be done safely (as judged by the investigator) during the same procedure in which the fresh tumor tissue biopsy is obtained. Evaluation of PD-L1 expression status will be done in real-time while the study is ongoing.

Part 2B: Safety Run-in for Additional Treatment Regimen of Durvalumab and Tremelimumab Combination Therapy

Immunotherapy-naïve subjects with advanced HCC who have progressed on, are intolerant to, or have refused sorafenib-based therapy will be enrolled in Part 2B. Approximately 6 to 12 subjects will be enrolled into Arm D to evaluate a single higher dose of tremelimumab in combination with durvalumab.

- Arm D: durvalumab and tremelimumab combination therapy (D 1500 mg + T 300 mg ×1 dose followed by D 1500 mg Q4W)

Subjects in Part 2B Arm D will receive durvalumab 1500 mg and tremelimumab 300 mg combination therapy for 1 dose followed by durvalumab 1500 mg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. A Part 2B safety evaluation will be performed once 6 safety evaluable subjects have completed 4 weeks of follow-up. A safety evaluable subject is defined as a subject who has received at least 1 dose of study drug and completed at least 4 weeks of follow-up or discontinued treatment prior to the completion of 4 weeks of follow-up due to an adverse event.

Part 3: Randomized Arms Evaluating Durvalumab and Tremelimumab Combination Therapy, Durvalumab Monotherapy, and Tremelimumab Monotherapy

Immunotherapy-naïve subjects with advanced HCC who have progressed on, are intolerant to, or have refused sorafenib-based therapy will be stratified based on viral status (uninfected, HCV, or HBV) and sorafenib/VEGFR TKI therapy (refusers or all others).

Subjects may not enroll into any arm of Part 3 until enrollment in Part 2A and 2B have been completed and safety evaluation on the 6 subjects in Part 2B is completed. Subjects in Part 3 will be randomized at a ratio of 2:1:2 into 1 of up to 3 treatment arms with approximately 64 subjects in each of the durvalumab monotherapy or combination treatment arms (B and D) and approximately 32 subjects in the tremelimumab monotherapy treatment arm (Arm C). No prerequisite number of subjects for viral status is set for any arm in Part 3.

- Arm A (recruitment closed following protocol amendment 5): durvalumab and tremelimumab combination therapy (D 1500 mg + T 75 mg Q4W ×4 doses followed by D 1500 mg Q4W)
- Arm B: durvalumab monotherapy (D 1500 mg Q4W)
- Arm C: tremelimumab monotherapy (T 750 mg Q4W ×7 doses, then Q12W)
- Arm D: durvalumab and tremelimumab combination therapy (D 1500 mg + T 300 mg ×1 dose followed by D 1500 mg Q4W)

Subjects in Part 3 Arm A (recruitment closed following protocol amendment 5) will receive durvalumab 1500 mg and tremelimumab 75 mg combination therapy Q4W for 4 doses followed by durvalumab 1500 mg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. Patients randomized to this arm before protocol amendment 5 can continue on assigned study treatment (provided the investigator and patient think it is in the best interest of the patient) until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met. If a patient has not completed or started all 4 doses of tremelimumab, the patient may either continue to complete the full schedule or continue with durvalumab monotherapy only.

Subjects in Part 3 Arm B will receive durvalumab 1500 mg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met.

Subject in Part 3 Arm C will receive tremelimumab 750 mg monotherapy Q4W for 7 doses followed by Q12W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met.

Subject in Part 3 Arm D will receive durvalumab 1500 mg and tremelimumab 300 mg combination therapy for 1 dose followed by durvalumab 1500 mg/kg monotherapy Q4W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met.

All subjects in Part 3 will be evaluated for safety and efficacy as specified in the protocol, and their disease status will be primarily analyzed according to RECIST v1.1. An interim analysis will occur after approximately 10 subjects per treatment arm in Part 3 have completed 4 weeks of follow-up as described in Protocol Section 4.8.7. All subjects will be followed for survival until the end of study as defined in Protocol Section 6.3. CCI [redacted]

CCI [redacted]

Part 4: Single Arm Evaluating Durvalumab with Bevacizumab in Combination.

Subjects with advanced HCC who have not received any prior systemic therapy can be enrolled into Part 4. Approximately 100 subjects will be enrolled into a single arm in Part 4 to evaluate the safety and efficacy of durvalumab in combination with bevacizumab:

- Durvalumab 1120 mg and bevacizumab 15 mg/kg combination therapy Q3W until confirmed PD or any other discontinuation criterion as described in Protocol Section 4.1.6 is met.

All subjects will be followed for survival until the end of study as defined in Protocol Section 6.3.

Retreatment option for patients in the durvalumab plus tremelimumab combination therapy arms

Subjects who complete the specified cycles of a durvalumab and tremelimumab combination therapy regimen (with clinical benefit per investigator judgment), but subsequently have evidence of PD during the durvalumab portion of the combination regimen, with or without

confirmation according to investigator assessments using RECIST v1.1, may restart treatment with the combination.

The retreatment option is available for the following study cohorts and combination treatment arms: Part 1, Part 2A Arm A, Part 2B Arm D, Part 3 Arm A and Part 3 Arm D.

Subjects enrolled in the durvalumab and tremelimumab combination therapy arms who meet retreatment option criteria will follow the same treatment guidelines followed during the original treatment period, including the same dose and frequency of treatments and the same schedule of assessments. Subjects who meet the criteria for retreatment for their respective treatment arm may only receive retreatment once. Crossover within the study is not permitted, except for patients in Part 3 Arm A, who can be retreated with the durvalumab 1500mg + tremelimumab 300mg ×1 dose, with prior approval from the AstraZeneca Study Physician.

Treatment through progression is permitted for subjects taking the durvalumab and bevacizumab combination in Part 4. However, only durvalumab can be continued through progression while bevacizumab should be discontinued at initial radiological progression. Re-treatment is not permitted for subjects in Part 4.

1.3 Number of Subjects

The study consists of 6 parts (1A, 1B, 2A, 2B, 3 and 4). Part 1A is a safety run-in with initial enrollment of 6 uninfected or HCV+ subjects and staggered enrollment of 6 to 12 additional subjects (3+3 dose-escalation design) with HBV+ disease as outlined in Section 1.2. The sample sizes for Part 1A were determined empirically and are consistent with those used in clinical studies to evaluate the safety of a proposed administered dose in a new patient population. There is a 47% to 91% probability of observing at least one AE from 6 subjects if the true incidence rate is 10% to 33%.

For Part 1B, approximately 24 subjects will be enrolled. For the combined sample size of approximately 36 subjects for Part 1A and Part 1B, there is an 84% to 98% probability of observing at least one AE if the true incidence rate is 5% to 10%. The sample size of 36 subjects is chosen to obtain a preliminary assessment of antitumor activity. Table 1.3-1 provides the estimated objective response rate (ORR) and the exact 95% confidence intervals (CI) for a range of possible responses out of 36 subjects. The preliminary assessment of antitumor activity will help determine if there is sufficient evidence of clinical efficacy to warrant continued enrollment of subjects in Part 2.

Table 1.3-1 Estimated Objective Response Rates and Exact 95% CI out of 36 Subjects

Number (%) of Responses	2 (5.6)	4 (11.1)	6 (16.7)	8 (22.2)	10 (27.8)	12 (33.3)	14 (38.9)
Lower limit of 95% CI	0.7%	3.1%	6.4%	10.1%	14.2%	18.6%	23.1%
Upper limit of 95% CI	18.7%	26.1%	32.8%	39.2%	45.2%	50.9%	56.5%

CI = confidence interval.

Part 2A dose-expansion analysis cohort will include approximately 108 subjects who will be globally recruited. Subjects will be stratified based on viral status (HBV vs HCV vs uninfected) and PD-L1 status (positive, negative, or non-evaluable). Within each stratum subjects will be randomly assigned in a 1:1:1 ratio to 3 treatment arms (Arms A, B, and C). With 12 subjects (per viral status cohort) treated with MED4736 or tremelimumab monotherapy respectively, there is a 72% to 86% probability of observing at least 1 AE from 12 subjects if the true incidence rate is 10% to 15%. All subjects recruited in the China-specific expansion of Part 2A after global recruitment of approximately 108 subjects has ended in Part 2A, will be analyzed separately.

Part 2B is a safety cohort for the safety evaluation of durvalumab in combination with a higher single-dose of tremelimumab (Arm D) in which approximately 6 to 12 subjects will be enrolled.

Part 3 is a dose-expansion cohort in which approximately 200 subjects will be randomized. Following protocol amendment 5 recruitment to Arm A was closed, and subjects in Part 3 will be assigned randomly at a ratio of 2:1:2 to 3 treatment arms (Arms B, C, and D) The stratification factors in the block randomization procedure are viral status (HBV, HCV, or uninfected) and sorafenib/VEGFR TKI therapy (refusers, or all others). With 64 subjects treated with durvalumab or tremelimumab monotherapy respectively, there is a 96% probability of observing at least 1 AE from 64 subjects if the true incidence rate is 5%.

Part 4 is a single arm cohort to evaluate safety and efficacy of durvalumab in combination with bevacizumab. The primary efficacy endpoint is ORR assessed by BICR according to RECIST 1.1. If the true ORR is 32%, a sample size of 100 subjects will provide at least 83% power to test the hypothesis that ORR is greater than 18.8% (Kudo et al 2018) at a 0.05 (2-sided) significance level. Other hypothesized ORRs and corresponding power values are given in Table 1.3-2.

Table 1.3-2 Power Calculations to Detect Difference Between True ORR versus Control ORR = 18.8%, n=100

True ORR	30%	31%	32%	33%	34%	35%
Power	70%	77%	83%	88%	92%	94%

Based on Exact Binomial Test, Alpha = 0.05 (2-sided)

With 100 patients, the precision of the estimation of ORR in the overall study population will be within:

• +/- 8% if the ORR is 10% (i.e. 95% CI 4.9%, 17.6%)
• +/- 9% if the ORR is 15% (i.e. 95% CI 8.6%, 23.5%)
• +/-11% if the ORR is 35% (i.e. 95% CI 25.7%, 45.2%)

There is a 99% probability of observing at least 1 AE from 100 subjects if the true incidence rate is 5%.

To evaluate the efficacy in terms of ORR of durvalumab and tremelimumab administered as monotherapy and combination therapy, data from Part 2 (Parts 2A and 2B) and Part 3 will be analyzed for each part separately and for both parts combined. If the true ORR is 20% in the durvalumab and tremelimumab combination therapy, a sample size of 100 subjects will provide at least 95% power to test the hypothesis that ORR is greater than 7% at a 0.05 significance level.

2 ANALYSIS SETS

2.1 Definition of Analysis Sets

The analysis sets are defined in Table 2.1-1. At interim analyses, analysis sets may include patients from only selected study Parts of interest.

Table 2.1-1 Analysis Populations

Population	Description
Full Analysis Set (FAS)	All randomized subjects (Parts 2A and 3) or subjects who were allocated to treatment (Parts 1A, 1B, 2B and 4), including subjects who were randomized in error. The FAS will be used for all efficacy analyses. Treatment arms will be compared on the basis of randomized/assigned study investigational product(s), regardless of the study investigational product(s) actually received. Subjects who were randomized/assigned to treatment but did not subsequently go on to receive study investigational product(s) will be included in the analysis in the treatment arm to which they were randomized/assigned. For non-randomized study parts, subjects are assigned to the treatment arm noted on exposure case report forms.
Response Evaluable Population	All treated subjects who have a baseline tumor assessment and measurable disease at baseline. Treatment arms will be compared on the basis of allocated or randomized study investigational product(s), regardless of the study investigational product(s) actually received. This population will be used only for ORR analysis.
Safety Analysis Set	Subjects who receive any study investigational product(s). If a subject receives more than one study investigational product, the subject will be summarized in the group based on the actual initial treatment (rather than the most total exposure) regardless of any subsequent treatment.

Table 2.1-1 Analysis Populations

Population	Description
CCI [redacted]	
DLT Evaluable Set	The DLT evaluable population includes all subjects enrolled in Part 1A who receive at least one dose of study investigational product and complete the safety follow-up through the DLT evaluation period or experience any DLT. Completing safety follow-up through the DLT evaluation period means being observed for at least 4 weeks (28 days) since the start of treatment (date of last available non-missing assessment – date of first dose + 1 >= 28 days).

2.2 Protocol Deviations

The following general categories will be considered important deviations and will be listed and discussed in the clinical study report (CSR) as appropriate for the study:

1. Patients randomized (Parts 2A and 3)/allocated to treatment (Parts 1, 2B and 4) but who did not receive study treatment.

2. Patients who deviate from key entry criteria per the Clinical Study Protocol (CSP) Amendment 6 (4-Jan-2019).

    a) Inclusion criteria: 5, 6.

    b) Exclusion criteria: 1, 2, 4, 8, 11, 23.

3. Baseline RECIST scan > 42 days before randomization (Parts 2A and 3) or first dose of study drug (Parts 1, 2B and 4).

4. No baseline RECIST 1.1 assessment on or before date of randomization (Parts 2A and 3) or first dose of study drug (Parts 1, 2B and 4).

5. Received prohibited concomitant systemic anticancer therapy. Please refer to the Protocol Section 4.7.2 for the systemic anticancer agents that are detailed as being ‘excluded’ from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to database lock.

6. Patients randomized/allocated to treatment who received their randomized/allocated study treatment at an incorrect dose or received an alternative study treatment to that which they were randomized/allocated.

7. Did not have the intended disease or indication:

    - Inclusion criteria 3 (per CSP Amendment 6 of 4-Jan-2019): Advanced HCC with diagnosis confirmed pathologically or with non-invasive methods.

    - Inclusion criteria 4 (per CSP Amendment 6 of 4-Jan-2019): Immunotherapy-naive and have either progressed on, are intolerant to, or refused treatment with sorafenib or another approved VEGFR TKI. (Part 4 only: Must not have received prior systemic therapy for HCC.)

Subjects who receive the wrong study treatment at any time will still be included in the Safety Analysis Set as described in Section 2.1. During the study, decisions on how to handle errors in treatment dispensing (with regards to continuation/discontinuation of study treatment or, if applicable, analytically) will be made on an individual basis with written instruction from the study team leader, physician, and/or statistician.

The important protocol deviations will be listed and summarized by randomized (Parts 2A and 3)/allocated treatment (Parts 1, 2B and 4). Deviation 1 will lead to exclusion from the Safety Analysis Set. None of the other deviations will lead to patients being excluded from the analysis sets described in Section 2.1 CCI [redacted]

If the deviations are serious enough to have the potential to impact the primary analysis, sensitivity analyses may be performed. The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of DBL and will be documented prior to the primary analysis being conducted.

3 STATISTICAL METHODS

3.1 General Considerations

For both efficacy and safety data, summaries and analyses will be performed by each part separately and for Part 2 and Part 3 combined. Dosing regimens in Arms A, B and C of Part 2A and Arm D of Part 2B are equivalent to the corresponding Arms A, B, C, and D in Part 3, and the subjects with the same dosing regimens will be combined accordingly for analyses. For instance, subjects from Part 2A Arm A and Part 3 Arm A will be grouped as Arm A of Parts 2 and 3 combined data.

All Part 1 study data will be summarized in the CSR under one treatment arm:

1) "Treme 1 mg/kg x4 doses + Durva 20 mg/kg" in tables, "T75+D" in figures.

All Part 4 study data will be summarized in the CSR under one treatment arm:

2) "Durva 1120 mg + Beva 15 mg/kg" in tables, "D1120+B15" in figures

The treatment arms in TFLs summarizing data from Part 2, Part 3, or Parts 2 and 3 combined will be displayed with these labels in the following order:

1) "Durva 1500 mg" in tables, "D" in figures
2) "Treme 300 mg ×1 dose + Durva 1500 mg" in tables, "T300+D" in figures
3) "Treme 750 mg" in tables, "T" in figures
4) "Treme 75 mg ×4 doses + Durva 1500 mg" in tables, "T75+D" in figures

Each TFL summarizing data from Part 2 or Parts 2 and 3 combined will have a footnote explaining weight-based dosing used in Part 2A.

The China cohort, who will be analyzed separately from Part 2 subjects, will include all subjects recruited in the China-specific expansion of Part 2A after global recruitment has ended in Part 2A. Data from subjects in the China cohort will be summarized by the following treatment arms:

1) "Durva 20 mg/kg mg" in tables, "D" in figures
2) "Treme 10 mg/kg" in tables, "T" in figures
3) "Treme 1 mg/kg x4 doses + Durva 20 mg/kg" in tables, "T75+D" in figures

Efficacy data will be summarized and analyzed using the Full Analysis Set. Safety and exposure data will be summarized and analyzed using the Safety Analysis Set. Study population and demography data will be summarized and analyzed using the Full Analysis Set.

Study data provided in data listings will be sorted by study part, treatment group and subject number. The following study data will be listed:

- Patients discontinued from study treatment
- Patients with important protocol deviations
- Demographic and baseline characteristics
- Concomitant medications on entry and during the study
- Patients randomized/allocated to treatment
- Administration of investigational products
- Individual efficacy response data
- Adverse events
- Adverse events occurring more than 90 days after discontinuation
- Individual laboratory assessments
- CCI [redacted]

CCI [redacted]

In general, descriptive statistics will be used for all variables, as appropriate. Categorical data will be summarized by the number and percentage of subjects falling within each category. Continuous variables will be summarized by descriptive statistics including the number of observations, mean, standard deviation, median, minimum, and maximum. Unless specified otherwise, percentages will be calculated out of the population total for the corresponding treatment arm. Overall totals will be calculated for baseline summaries only.

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. P-values should be displayed with 4 decimal places, confidence intervals with the same precision as the corresponding statistic, and all ratios with 1 decimal place.

All available data will be used and thus, missing data will not be imputed. Subjects with missing data for a parameter will be excluded from the summary of this parameter.

Missing safety data will generally not be imputed either. 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. Note that 0 should not be used as an imputed value in case the endpoint requires a log transformation. Additionally, adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

For immune-mediated adverse event summaries, an AE with outcome of unknown will be imputed as not resolved.

Furthermore:

- For missing diagnostic dates, if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.

- For missing AE/Concomitant Medications start dates, the following will be applied

    a. Missing day - Impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date

    b. Missing day and month - Impute 1st January unless year is the same as first dose date then impute first dose date

    c. Completely missing - impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date

- For missing AE/ Concomitant Medications end dates, the following will be applied:

    a. Missing day - Impute the last day of the month

    b. Missing day and month – impute 31st December

Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.

- If a patient is known to have died where only a partial death date is available then the date of death will be imputed as the latest of the last date known to be alive +1 from the database and the death date using the available information provided:

    a. For Missing day only – using the 1st of the month

    b. For Missing day and Month – using the 1st of January

Imputed death date = max(last date known to be alive +1, imputed partial death date).

If there is evidence of death but the date is entirely missing, it will be treated as missing, i.e. censored at the last known alive date.

Data analyses will be conducted using the SAS® System version 9.3 or higher (SAS Institute Inc., Cary, NC). The analytical results generated from SAS programs will be validated according to AstraZeneca SAS programming standards and AstraZeneca SAS validation procedures.

Unless otherwise specified, date of initiation of the first subsequent anticancer therapy should be the date of the first subsequent systemic anticancer therapy.

3.1.1 Baseline

In general, baseline is defined as the last non-missing measurement prior to (before or on) the first dose of study medication (per CSP, drug administration is the last procedure performed for a subject on a certain day). If there is more than one visit equally eligible to assess subject 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) with missing assessment time, the average can be taken as baseline value. In tables based on the Full Analysis Set (demographics, baseline characteristics) baseline should be calculated as the last value prior to enrollment (Parts 1, 2B, 4) or randomization (Parts 2A, 3) for patients who never received study treatment to allow summary for Full Analysis Set.

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. If no value exists before the first dose/administration, then the baseline value will be treated as missing. In case both date and time of the measurement are available, they should be compared with date and time of the first dose of study medication (e.g. laboratory, pharmacokinetics).

For analyses related to tumor assessment based on RECIST1.1 data, baseline is defined as the last non-missing assessment prior to (before or on) randomization date for Parts 2A and 3, or prior to (before or on) first dose date for Parts 1, 2B and 4.

In all summaries of change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The % change from baseline will be calculated as:

(post-baseline value - baseline value) / baseline value × 100.

Study day will be calculated as:

Date of assessment – Date of first dose of study medication +1.

3.1.2 Visit windows for safety assessments

Safety data including laboratory test results, ECG, vital signs, and ECOG performance status will be summarized descriptively by the visit at scheduled time of evaluation.

Time windows will need defining for any presentations that summarize safety data values by visit. The following conventions should also 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.

Parts 1-3

Visit windows will be defined as follows in study Parts 1-3 (schedule of study procedures Q2W for Week 1-25, and Q4W from Week 25):

- Baseline, visit window Low-D1

- Week 2, Day 15 (2×7+1), visits window D2 - D21
- Week 4, Day 29 (4×7+1), visit window D22 - D35
- Week 6, Day 43 (6×7+1), visit window D36 - D49
- Week 8, Day 57 (8×7+1), visit window D50 - D63
- Week 10, Day 71 (10×7+1), visit window D64 - D77
- Week 12, Day 85 (12×7+1), visit window D78 - D91
- Week 14, Day 99 (14×7+1), visit window D92 – D105
- Week 16, Day 113 (16×7+1), visit window D106 – D119
- Week 18, Day 127 (18×7+1), visit window D120 – D133
- Week 20, Day 141 (20×7+1), visit window D134 – D147
- Week 22, Day 155 (22×7+1), visit window D148 – D161
- Week 24, Day 169 (24×7+1), visit window D162 – D182
- Week 28, Day 197 (18×7+1), visit window D183 – D210 (Q4W from Week 29)
- Week 32, Day 225 (32×7+1), visit window D211 – D238
- Week 36, Day 253 (36×7+1), visit window D239 – D266
- … (and continued every 4 weeks until PD or discontinue from the study)
- Follow-up visit

Some assessments in Parts 1-3 are collected Q4W (thyroid function tests, urinalysis, CCI [redacted] or Q8W (ECOG).

For assessments collected Q4W during the whole study in Parts 1-3, the windows are:

- Baseline, visit window Low-D1
- Week 4, Day 29 (4×7+1), visit window D2 – D42
- Week 8, Day 57 (8×7+1), visit window D43 – D70
- Week 12, Day 85 (12×7+1), visit window D71 – D98
- Week 16, Day 113 (16×7+1), visit window D99 – D126
- … (and continued every 4 weeks until PD or discontinue from the study)
- Follow-up visit

For assessments collected Q8W during the whole study in Parts 1-3, the windows are:

- Baseline, visit window Low-D1
- Week 8, Day 57 (8×7+1), visit window D2 – D84
- Week 16, Day 113 (16×7+1), visit window D85 – D140
- Week 24, Day 169 (24×7+1), visit window D141 – D196
- Week 32, Day 225 (32×7+1), visit window D197 – D252
- … (and continued every 8 weeks until PD or discontinue from the study)

- Follow-up visit

Part 4

Visit windows will be defined as follows in study Part 4 (schedule of study procedures Q3W during the whole study):

- Baseline, visit window Low-D1
- Week 3, Day 22 (3×7+1), visits window D2 - D32
- Week 6, Day 43 (6×7+1), visit window D33 - D53
- Week 9, Day 64 (9×7+1), visit window D54 - D74
- Week 12, Day 85 (12×7+1), visit window D75 - D95
- … (and continued every 3 weeks until PD or discontinue from the study)
- Follow-up visit

Some assessments in Part 4 are collected Q6W (ECOG).

For assessments collected Q6W during the whole study in Part 4, the windows are:

- Baseline, visit window Low-D1
- Week 6, Day 43 (6×7+1), visit window D2 – D63
- Week 12, Day 85 (12×7+1), visit window D64 – D105
- Week 18, Day 127 (18×7+1), visit window D106 – D147
- Week 24, Day 169 (24×7+1), visit window D148 – D189
- … (and continued every 6 weeks until PD or discontinue from the study)
- Follow-up visit

Listings will display all values contributing to a time point for a subject. For summary, 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 the patient that contributed to the summary table, wherever feasible.

For summaries at a subject 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. Where safety data are summarized over time, study day will be calculated in relation to date of first treatment.

CCI

3.1.4 Analysis of Chinese Subjects in Part 2A

The China cohort will include all subjects recruited in the China-specific expansion of Part 2A after global recruitment has ended in Part 2A. Up to 60 additional Chinese subjects will be enrolled in Part 2A (approximately 20 per arm; treatment arms A, B, and C). The safety and efficacy data from subjects in the China cohort will be analyzed separately from Part 2 subjects. These analyses will use the same endpoint definitions and the same statistical methods as detailed in Section 2.

The data cut-off for China cohort analyses will be the same as the global study data cut-off. CCI

3.2 Demographic and baseline characteristics, exposure and concomitant medications

3.2.1 Subject Disposition and Completion Status

A summary of subjects treated will be provided. In addition, summaries of number and percent of subjects that completed or discontinued treatment and the reasons for discontinuation during the initial treatment (Full Analysis Set) will be provided.

Summaries using the number and percent of subject status at the end of study and the reasons for ending study will be provided.

A summary of the number of patients who have died will be produced. This will be presented by treatment arm and include the cause of death (toxicity related to investigation product or disease under investigation)

3.2.2 Demographics and Baseline Characteristics

The following will be summarized using Full Analysis Set:

- Demographics: age [age continuous (years), age group (<65, >=65- <75, >=75 years)], sex, ethnicity, race (if there is more than one race category marked for a patient, this patient should be counted under an additionally created race category "Mixed");
- Baseline patient characteristics: height (cm), weight [weight continuous (kg), weight group (<70, >=70- <90, >=90 kg)], body mass index (BMI) [BMI continuous (kg/m2), BMI group (<18.5, ≥18.5 – 25.0, ≥25.0 – <30.0, ≥30.0 kg/m^2)], PD-L1 status (for the definition please see Section 3.3.3.10), ECOG performance status, prior treatment with sorafenib/VEGFR TKI (Yes, No), and Child-Pugh Class/Score;
- Patient recruitment by country and center;
- Stratification factors by IWRS and CRF:
    - Viral status (HBV, HCV, uninfected) and PD-L1 expression (positive, negative, non-evaluable) for Part 2A;
    - Viral status (HBV, HCV, uninfected) and sorafenib/VEGFR TKI treatment (refusers, all others) for Part 3;
- Previous anticancer systemic therapy;
- Previous anticancer radiation therapy;

- Prior cancer related surgery;
- Medical history;
- Disease characteristics and initial diagnosis: Disease stage, tumor stage, node stage, metastasis stage, BCLC stage, CLIP score, time from diagnosis to first dose (months), and time from diagnosis to randomization (months) for Part 2A and Part 3;
- Disease characteristics at screening: Disease stage, BCLC stage, CLIP score, CCI [redacted] see Section 3.3.3.11), sorafenib therapy, CCI [redacted] Macrovascular invasion, and extrahepatic spread;

The following tables should be repeated for the subset of subjects (if any) experiencing retreatment: patient disposition (discontinuation reason for retreatment period), demographics, patient characteristics at baseline, disease characteristics at initial diagnosis, disease characteristics at screening, previous cancer treatment, TNM classification at screening, and subsequent cancer therapy.

3.2.3 Concomitant and other treatments

Medications received prior to, concomitantly, or post-treatment will be coded using the AstraZeneca Drug Dictionary Anatomical Therapeutic Chemical (ATC) Classification codes. Concomitant medications will be summarized for the FAS by ATC classification codes.

For the purpose of inclusion in prior and/or concomitant mediation or therapy summaries, incomplete medication or start and stop dates will be imputed like for AEs as detailed in Section 3.1.

Prior medications, concomitant and post-randomized treatment medications are defined based on imputed start and stop dates as follows:

- Prior medications are those taken prior to study treatment with a stop date prior to the first dose of study treatment.
- Concomitant medications are those with a stop date on or after the first dose date of study treatment (and could have started prior to or during treatment).
- Post-treatment medications are those with a start date after the last dose date of study treatment.

In addition, all post-treatment anticancer therapies will be summarized for the full analysis set.

The following summaries will be produced:

- Summary of prior medications
- Summary of concomitant medications
- Summary of post study treatment cancer therapies

All concomitant and other treatment data will be listed. Missing coding terms should be listed and summarised as "Not coded".

The WHO Drug dictionary will be used for concomitant medications coding.

Concomitant medications will be split into allowed and disallowed categories for table summaries. Classification of allowed and disallowed medications will be reviewed by the Study Team Physician prior to database lock. Disallowed medications will be identified, and all medications which are not disallowed, will be reported as allowed medications.

3.2.4 Study Drug Exposure and Dose Intensity

3.2.4.1 Study drug exposure

Exposure to study drug(s) (intended and actual), time on study, cycle delays (all arms), and infusion interruptions (all arms) will be summarized. For the duration of exposure "total treatment years" will be calculated next to the basic summary statistics.

Exposure will be defined as follows.

Total (or intended) exposure of study medications in Parts 1, 2 and 3:

a) Durvalumab

- Total (or intended) durvalumab exposure = min (last durvalumab dose date where dose>0 + 27 days, date of death, date of DCO) – first dose date +1

b) Tremelimumab

For Part 1, Part 2A Arms A and B, Part 2B Arm D, and Part 3 Arms A, B and D:

- Total (or intended) tremelimumab exposure = min (last tremelimumab dose date where dose>0 + 27 days, date of death, date of DCO) – first dose date +1

For Part 2A Arm C and Part 3 Arm C:

- Total (or intended) tremelimumab exposure = min (last tremelimumab dose date where dose>0 + 27 days, date of death, date of DCO) – first dose date +1 for the first 7 doses,

and

for the remaining doses:

- Total (or intended) tremelimumab exposure = min (last tremelimumab dose date where dose>0 + 83 days, date of death, date of DCO) – first dose date +1

Total (or intended) exposure of study medications in Part 4:

- Total (or intended) durvalumab exposure = min (last durvalumab dose date where dose>0 + 20 days, date of death, date of DCO) – first dose date +1

- Total (or intended) bevacizumab exposure = min (last bevacizumab dose date where dose>0 + 20 days, date of death, date of DCO) – first dose date +1

Actual exposure:

- Actual exposure = intended exposure – total duration of cycle delays/interruptions, where intended exposure will be calculated as above, and a cycle delay/dose interruption is defined as any length of time where the patient has not taken any of the planned daily dose.

Total duration of cycle delays will be summarized (definitions below).

Durvalumab and tremelimumab in Parts 1, 2 and 3:

Since patients will receive drug via IV infusions every 4 weeks (Q4W) until confirmed PD the duration of cycle delay will be calculated as:

- Total duration of dose delays = max (0, Sum of (Date of the dose start - Date of previous dose end – 28 days)),

with the exception of Part 2A Arm C and Part 3 Arm C for which the patients receive 7 doses of drug every 4 weeks (Q4W) and then every 12 weeks (Q12W) until confirmed PD, the duration of cycle delay will be calculated as:

- Total duration of dose delays = max (0, Sum of (Date of the dose start - Date of previous dose end – 28 days)) for the first 7 doses and for the remaining doses:

- Total duration of dose delays = max (0, Sum of (Date of the dose start - Date of previous dose end – 84 days)).

Durvalumab and bevacizumab in Part 4:

Since patient will receive drug via IV infusions every 3 weeks (Q3W) until confirmed PD the duration of a cycle delay will be calculated as:

- Total duration of dose delays = max (0, Sum of (Date of the dose start - Date of previous dose end – 21 days)).

Exposure should be calculated separately for the following three periods: initial treatment, retreatment, and total study. For durvalumab monotherapy and tremelimumab monotherapy arms, initial treatment and total study will be the same. For combination therapy arms, initial treatment includes data before retreatment, retreatment includes data from retreatment onwards, and total study includes all data from initial treatment and retreatment.

Time on study should be defined as follows (all treatment arms):

- Time on study (days) = last available non-missing assessment date – randomization date +1.

Total exposure, actual exposure and time on study expressed in months will be summarized in tables as continuous data as well as categorical data. The duration in months will be calculated as follows:

- Duration (month) = Duration in days / (365.25/12)

3.2.4.2 Dose intensity

Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation. It should be calculated for each study drug separately.

RDI will be defined as follows:

> RDI = 100% * d/D, where d is the actual cumulative dose delivered up the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule.

Intended cumulative dose will be calculated as follows:

- Intended number of cycles * intended dose per cycle;
- Number of intended cycles in Part 1, 2, 3 = floor((last dose date – first dose date + 3)/28) +1.
    - For Treme mono arm patients with exposure <= 28 weeks (last dose date – first dose date+1 <=28*7 + 3 days; corresponding to the first 7 doses period):
      Number of intended cycles = floor((last dose date – first dose date + 3)/28) +1.
    - For Treme mono arm patients with exposure > 28 weeks (expected at least 8 doses; last dose date – first dose date+1 > 4*7 + 3 days):
      Number of intended cycles = 7 + floor((last dose date – first dose date + 3 – 28*7)/84) +1.
- Number of intended cycles in Part 4 = floor((last dose date – first dose date + 3)/21) +1.

Intended dose during treatment and retreatment period will be assigned as follows:

Part 1:

- durvalumab 20 mg/kg, tremelimumab 1 mg/kg;

Part 2:

- Arm A: durvalumab 20 mg/kg, tremelimumab 1 mg/kg;
- Arm B: durvalumab 20 mg/kg (retreatment not possible);
- Arm C: tremelimumab 10mg/kg (retreatment not possible);
- Arm D: durvalumab 1500 mg (20 mg/kg if weight ⩽30 kg), tremelimumab 300 mg (weight-based dosing not applicable)

Part 3:

- Arm A: 1500 mg durvalumab (20 mg/kg if weight ⩽30 kg), 75 mg x4 tremelimumab (1 mg/kg if weight ⩽30 kg);
- Arm B: 1500 mg durvalumab (20 mg/kg if weight ⩽30 kg) (retreatment not possible);
- Arm C: tremelimumab 750 mg (10 mg/kg if weight ⩽30 kg) (retreatment not possible);
- Arm D: 1500 mg durvalumab (20 mg/kg if weight ⩽30 kg), 300 mg tremelimumab (weight-based dosing not applicable);

Part 4:

- durvalumab 1120 mg (15 mg/kg if weight ⩽ 30 kg) and bevacizumab 15 mg/kg (retreatment not possible);

For weight-based dosing treatment arms, when calculating the intended dose for visits not completed by a patient, the intended dose from the last available visit should be used.

Dose intensity will be summarized using the Safety Analysis Set. For Parts 1-3 summaries should be done for three study periods: initial treatment, retreatment (if applicable), and total study. For Part 4 only the overall total study summaries will be shown, since retreatment is not allowed in Part 4.

3.3 Efficacy Analyses

Analyses of response-related endpoints and the corresponding time-to-event endpoints will be based on an application of RECIST 1.1 guidelines (Eisenhauer et al, 2009) according to Investigator-assessed tumor measurements and BICR-assessed tumor measurements. In addition, both unconfirmed and confirmed responses will be summarized and analyzed utilizing Investigator assessments and BICR assessments.

For randomized study parts (Parts 2A and Part 3) treatment arms will be compared on the basis of randomized study investigational product(s), regardless of the study investigational product(s) actually received.

3.3.1 Efficacy Endpoints

The efficacy endpoints include best objective response (BoR), objective response rate (ORR), disease control rate (DCR), time to response (TTR), duration of response (DoR), time to progression (TTP), progression-free survival (PFS), overall survival (OS) and landmark 12-month OS (OS-12m), 18-month OS (OS-18m) and 24-month OS (OS-24m).

3.3.1.1 Best objective response

Best objective response (BoR) is defined as the best response (in the order of CR, PR, SD, PD, and not evaluable) among all overall responses recorded from randomization (Parts 2A, Part 3) or the start of treatment (Parts 1, 2B and 4) until progression (per RECIST 1.1), or the last evaluable disease assessment in the absence of PD prior to the initiation of subsequent anticancer therapy (date of the first subsequent systemic therapy), whichever occurs first. BoR will be based on all post-baseline disease assessments that occur prior to the initiation of alternative anticancer therapy.

The best objective response of CR or PR must be confirmed by repeat imaging not less than 4 weeks (28 days) after the visit when the response was first observed. A confirmed response of CR/PR means that a response of CR/PR is recorded at 1 visit and confirmed by repeat imaging 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 visit.

In the case where a patient has two non-consecutive visit responses of PR, then, as long as the time between the 2 visits of PR is greater than 4 weeks and there is no PD between the PR visits, the patient will be defined as a responder. Similarly, if a patient has visit responses of CR, NE, CR, then, as long as the time between the 2 visits of CR is greater than 4 weeks, then a best response of CR will be assigned.

In order to assign a best objective response of SD (unconfirmed CR/PR will be considered SD):

- For Parts 1-3, SD should be recorded at least 8 weeks minus 1 week, i.e. at least 49 days (to allow for an early assessment within the assessment window), after first dose of treatment with investigational product (Parts 1, 2B), or after randomization date (Parts 2A and Part 3).

- For Part 4, SD should be recorded at least 9 weeks minus 1 week, i.e. at least 56 days (to allow for an early assessment within the assessment window), after first dose of treatment with investigational product.

The analysis will be based on the Full Analysis Set.

3.3.1.2 Objective response rate

Objective response rate (ORR) is defined as the rate of best objective response (BoR) of confirmed CR/PR according to RECIST v1.1. Refer to Section 3.3.1.1 for confirmed CR/PR. Subjects who go off treatment without progression, receive a subsequent anticancer therapy (note that for this analysis radiotherapy is not considered a subsequent anticancer therapy), and then respond will not be included as responders in the ORR. Subjects who are missing overall response assessments will be considered non-responders and counted in the denominator of the ORR. The analysis will be based on the Full Analysis Set and on the Response Evaluable Population. The unconfirmed ORR will be calculated additionally and reported in CSR.

3.3.1.3 Disease control rate

Disease control rate (DCR) is defined as the proportion of subjects with a BoR of confirmed CR/PR, or SD based on the Full Analysis Set (unconfirmed CR/PR will be considered a SD).

3.3.1.4 Time to response

Time to response (TTR) is defined as the time from randomization (Parts 2A and 3) or from first dose of study treatment (Parts 1, 2B and 4) to the first documentation of an objective response. Only subjects who have achieved a confirmed objective response (confirmed CR/PR) will be evaluated for TTR. For the definition of a confirmed response please see Section 3.3.1.1.

- TTR = Date of response – Date of randomization/first dose of investigational product + 1.

3.3.1.5 Progression-free survival

Progression-free survival (PFS) is defined as the time from date of randomization (Parts 2A and 3) or date of first dose of study drug (Parts 1, 2B and 4) to the first documented radiographic disease progression (per RECIST v1.1) or death due to any cause, whichever occurs first, regardless of whether the patient withdraws from therapy or receives another anticancer 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 patient progresses or dies after 2 or more consecutive missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the 2 missed visits (Note: An NE visit is not considered a missed visit).

Given the different scheduled RECIST visit assessment scheme between Part 1-3 (Q8W) and Part 4 (Q9W) the definition of 2 missed visits (both scheduled and unscheduled) will differ

between these study parts. For Parts 1-3 two missing visits will equate to 18 weeks (126 days) since the previous RECIST assessment (i.e. 2 × 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks). For Part 4 two missing visits will equate to 20 weeks (140 days) since the previous RECIST assessment (i.e. 2 × 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks).

If the patient has no evaluable visits or does not have baseline data, they will be censored at Day 1 (randomization day for Parts 2A and 3, and first dose date for Parts 1, 2B and 4) unless they die within 2 visits of baseline (2 × 8 weeks + 1 week = 119 days for Parts 1-3, and 2 × 9 weeks + 1 week = 133 days for Part 4, since there is no need to account for an early initial visit in this case), then they will be treated as an event with date of death as the event date (see Table 3.3-1).

Table 3.3-1 Summary of censoring guidelines for PFS

Assessment	Date of PD/Death or Censoring	PFS Outcome
Progression documented between scheduled visits	Date of assessment of progression	Event
No baseline or evaluable tumor assessments and death within 2 visits (119 days for Parts 1-3,133 days for Part 4) of baseline (or of Day 1 if baseline is not available)	Date of death	Event
Death between assessment visits	Date of death	Event
No baseline assessments or no evaluable response visits (excluding deaths within 2 visits of baseline)	Randomization/ first dose date	Censored
No PD or death at time of analysis or lost to follow-up	Date of last evaluable tumor assessment	Censored
Death or progression after 2 or more missed visits (126 days for Parts 1-3, 140 days for Part 4)	Date of last evaluable tumor assessment prior to the 2 missed visits	Censored

The date of radiographic PD (per RECIST v1.1) will be the first date at which any objective diagnostic test provides data indicating PD. Specifically, the date of PD will be the earliest of the dates of components that trigger the PD.

3.3.1.6 Duration of response

DoR will be defined as the time from the date of first documented response (which is subsequently confirmed as described in Section 3.3.1.1) until date of documented progression or death in the absence of disease progression (i.e. date of PFS event or censoring – date of first response + 1). 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 that was CR or PR that was subsequently confirmed. Only subjects with a BoR of confirmed complete response or confirmed partial response will be included in the analysis of DoR.

3.3.1.7 Time to progression

Time to progression (TTP) is defined as the time from the date of randomization (Parts 2A and 3) or date of first dosing (Parts 1, 2B and 4) to the first date of documented radiographic PD. TTP is defined as per PFS however if subjects died without tumor progression, they will be censored at the time of death.

TTP is calculated as:

Date of PD or censoring – Date of randomization/first dose of study drug + 1.

3.3.1.8 Overall survival

Overall survival (OS) is defined as the time from date of randomization (Parts 2A and 3) or date of first dose of investigational product (Parts 1, 2B and 4) to death due to any cause regardless of whether the patient withdraws from randomized therapy or receives another anticancer therapy. If there is no death reported for a subject by the data cutoff date for overall survival analysis, OS will be censored at the last known alive date. If the last known date alive is after the data cutoff date for the OS analysis, the last known date alive will be truncated at the date cutoff date.

Overall survival (OS) is calculated in months as:

- Date of death or censoring – Date of randomization (Parts 2A and 3) or date of first dose of investigational product (Parts 1, 2B and 4) + 1.

For any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it will 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 if a survival sweep is not performed). The last date for each individual patient is defined as the latest among the following dates recorded on the case report forms (CRFs):

• AE start and stop dates (Adverse Events, Serious Adverse Events Report), AE change date (Adverse Events)

• Admission and discharge dates of hospitalization (Serious Adverse Event Report)

• Date of last dose of study treatment (study drug exposure CRF forms)

• Laboratory test dates (Chemistry, Thyroid Function Test, CCI [redacted] Hematology, Coagulation, Urinalysis, Virology)

• Disease assessment dates on RECIST CRF (Target Lesions, Non-target Lesions, New Lesions, including retreatment pages)

• Start and stop dates of alternative anticancer treatment (Radiation therapy, Surgical and Medical procedures, Subsequent systemic therapy, Subsequent radiation treatment, Subsequent cancer related surgery)

• Date last known alive on survival status CRF (Survival status)

• End of study date (End of study: Date of study completion/ withdrawal/ Date of last contact/ Date of other)

Missing or incomplete death dates will be imputed as described in Section 3.1.

The landmark 1-year OS is defined as the proportion of patients alive at one year from the date of randomization (Parts 2A and 3) or date of first dose of investigational product (Parts 1, 2B and 4). OS-18m and OS-24m are defined similarly.

CCI [redacted]

For analysis, OS in days should be converted to OS in months as follows:

OS (months) = OS (days) / (365.25/12)

3.3.2 Derivation of RECIST Visit Responses

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

RECIST 1.1 assessments will be performed on images from CT (preferred) or MRI each preferably with IV contrast of the chest, abdomen (including liver and adrenal glands), and pelvis. Pelvic imaging is recommended only when primary or metastatic disease in the pelvic region is likely. Additional anatomy should be imaged based on signs and symptoms of individual patients at baseline and follow-up.

Baseline radiological tumor assessments are to be performed before or on the date of randomization (Parts 2A and 3) or date of first dose of study drug (Parts 1, 2B and 4) and, ideally, as close as possible to the start of study drug(s) (CSP Tables 4.2.1-1, 4.2.2-1, 4.2.2-2 and 4.2.3-1).

Follow-up assessments will be performed every 8 weeks (± 7 days) for study Parts 1-3 and every 9 weeks in Part 4, as indicated in the schedule of procedures presented in the protocol (Tables 4.2.1-1, 4.2.2-1, 4.2.2-2, 4.2.2-3, 4.2.2-4 and 4.2.3-1) until disease progression. The imaging schedule must be followed regardless of any delays in dosing. The subsequent (confirmatory)

scans should preferably be performed at the next scheduled visit (relative to the date of randomization) and no less than 4 weeks after the initial assessment of progressive disease (PD) (in the absence of clinically significant deterioration). Additional scans post confirmed PD will be completed per standard clinical practice.

If an unscheduled assessment was performed and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their next scheduled visits (relative to randomization/allocation to treatment). 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. All confirmatory scans should be recorded on the database.

Confirmation of Progression

In clinically stable patients who continue to receive treatment beyond an initial RECIST 1.1-defined PD, the immediate prior radiological progression would be considered confirmed if the following criteria are applied to the subsequent scan (acquired preferably at the next scheduled visit relative to the date of randomization (Parts 2A and 3) or first dose date (Parts 1, 2B and 4) but no less than 4 weeks after the prior RECIST 1.1-defined assessment of PD):

- ≥20% increase in the sum of diameters of target lesions (TL) compared with the nadir at 2 consecutive visits with an absolute increase of 5 mm.
- And/or significant progression (worsening) of non-target lesions (NTL) or of pre-existing new lesions at the subsequent time point.
- And/or additional (brand) new unequivocal lesions at the subsequent time point.

Note: A confirmation of progression does not affect imaging endpoints that use progression, since progression is defined by RECIST 1.1 as the first date of progression (with or without confirmation by later scans).

In the absence of significant clinical deterioration, the investigator may continue treatment until radiological progression is confirmed. If progression is not confirmed, then the patient should continue treatment and on-treatment assessments. A subsequent follow-up scan is required for all patients after a RECIST 1.1-defined PD, even if a subsequent treatment is started.

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

3.3.2.1 Site investigator assessment using RECIST 1.1

All RECIST assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues study treatment or receives another anticancer therapy. All patients (all cohorts) will be assessed using images collected Q8W ±7 days (Parts 1-3) or Q9W ±7 days (Part 4) relative to the date of randomization/allocation to treatment until confirmed objective disease progression. Additional scans post confirmed PD will be completed per standard clinical practice for subjects who discontinue treatment.

From the investigator's review of the imaging scans, the RECIST tumor response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous assessments. Baseline will be assessed before or on randomization date (Parts 2A and 3) and first dose date (Parts 1, 2B and 4). 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 objective disease progression according to RECIST 1.1 in which case the response will be assigned as PD). Imaging and procedures performed before signing the ICF may be used for screening purposes if the patient consents.

RECIST outcomes (i.e. TTP, PFS, ORR etc.) will be calculated programmatically for the site investigator data (see Section 3.3.2.4) from the overall visit responses.

3.3.2.2 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 computed tomography (CT) or magnetic resonance imaging (MRI) and which is suitable for accurate repeated measurements.

A patient 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). Lymph nodes are collectively considered as a single organ (regardless of designation of 'local/regional' or 'distant'). Bilateral organs or multi-lobular organs are each considered as a single organ. If more than one baseline scan is recorded, then measurements from the one that is closest to and prior to the date of randomization (Parts 2A and 3)/ first dose (Parts 1, 2B and 4) 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 NTL at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

Measurable disease (i.e. at least one TL) is one of the entry criteria for the study. However, if a patient with non-measurable disease is enrolled in the study (i.e. no TLs but have non-measurable disease), the evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see Section 3.3.2.4 for further details). If a patient does not have measurable disease at baseline, then the TL visit response will be not applicable (NA).

Table 3.3-2 TL Visit Responses (RECIST 1.1)

Visit Responses	Description
Complete Response (CR)	Disappearance of all target lesions since baseline. Any pathological lymph nodes selected as target lesions must have a reduction in short axis to <10 mm.
Partial Response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive Disease (PD)	At least a 20% increase in the sum of diameters of target lesions, 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 mm.
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
Not Evaluable (NE)	Only relevant in certain situations (i.e. if any of the target lesions were not assessed or not evaluable or had a lesion intervention at this visit; and scaling up could not be performed for lesions with interventions). Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a target lesion response
Not Applicable (NA)	No target lesions are recorded at baseline

Rounding of TL data

The RECIST eCRF pages collect TL measurements as integers in millimeters. For calculation of PD and PR for TLs, percentage changes from baseline and previous minimum should be

rounded to 1 decimal place 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 at least a 20% increase, and an absolute increase of ≥ 5mm, from nadir.

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

Lymph nodes

For lymph nodes, if the short axis diameter reduces to < 10 mm 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 < 10 mm and all other TLs are 0 mm then although the sum may be >0 mm 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. 0 mm for non-nodal TLs or < 10 mm for lymph node TLs) 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 longest diameter increases by 20% but remains < 10 mm.

Step 2: If some lesion measurements are assigned "NE" but all other lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to NE irrespective of whether when referencing the sum of TL diameters the criteria for PD is 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 large 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. 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 5 mm 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.

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. They can only be identified as TLs in they are they are the only ones available.

Any TL (including lymph nodes), which has had intervention during the study (for example, radiotherapy / 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 scale up as described below, as long as there remain ≤ 1/3 of the TLs with missing measurements. If the scaling results in a visit response of PD then the patient 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 with interventions), 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 has a value of 0 recorded.

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 where appropriate (as per step 2 above).

Scaling (applicable only for lesion intervention)

If > 1/3 of target lesion measurements are treated as missing (because of intervention) then target lesion response will be NE, unless the sum of diameters of non-missing target lesion 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 target lesions has increased by 5 mm from nadir).

If ≤ 1/3 of the target lesion measurements are treated as 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).

The scaling may be applied at the time of the analysis and is not performed by the Investigator.

Example of scaling

Table 3.3-3 shows an example of TL scaling.

Table 3.3-3 Example of scaling

Lesion	Longest diameter at nadir visit (mm)	Longest diameter at follow-up visit (mm)
1	7	7
2	6	4
3	4	4
4	8	8
5	2	3 (Intervention)
Sum	**27**	**26**

Lesion 5 has had an intervention at the follow-up visit. The sum of lesions 1-4 at the follow-up is 23 mm. The sum of the corresponding lesions (1-4) at baseline visit is 25 mm.

Scale up as follows to give an estimated TL sum of 24.84 mm:

(23 / 25) * 27= 24.84 mm

Lesions that split in two or more parts

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

Lesions that merge

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

Change in method of assessment of TLs

CT and MRI are the only methods of assessment that can be used within the trial. 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.

3.3.2.3 Non-target lesions (NTLs) and new lesions

At each visit, the investigator should record an overall assessment of the NTL response. 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:

Table 3.3-4 NTL visit responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
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.
Non CR/Non PD	Persistence of one or more NTLs-with no evidence of progression.

Table 3.3-4 NTL visit responses

Visit Responses	Description
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.
Not Applicable (NA)	Only relevant if there are no NTLs at baseline

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 discontinuation of therapy. 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.

If a new lesion is equivocal, for example because of its small size, the treatment and tumor assessments should be continued until the new lesion has been confirmed. If repeat scans confirm there is a new lesion, then the progression date should be declared using the date of the initial scan.

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

A new lesion indicates a 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 and should be treated as NE in the derivation of overall visit response.

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 radiological evidence of disease progression at that time should continue to undergo tumor assessments where possible until radiological progression is observed.

3.3.2.4 Overall visit response

Table 3.3-5 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 3.3-5 Overall Visit Responses

Target Lesions	Non-target lesions	New Lesions	Overall Response
CR	CR or NA	No (or NE)	CR
NA	CR	No (or NE)	CR
CR	Non CR/Non PD or NE	No (or NE)	PR
PR	Non PD or NE or NA	No (or NE)	PR
SD	Non PD or NE or NA	No (or NE)	SD
NA	Non CR/Non PD	No (or NE)	SD
NE	Non PD or NE or NA	No (or NE)	NE
NA	NE	No (or NE)	NE
PD	Any	Any	PD
Any	PD	Any	PD
Any	Any	Yes	PD
NA	NA	No (or NE)	NED

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 TL/NTL at baseline), NED No evidence of disease.

3.3.2.5 Best objective response (BoR)

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in Section 3.3.2.4. It is the best response a patient has had following first dose of study medication (Parts 1, 2B and 4) or randomization (Parts 2A and 3), but prior to starting any subsequent cancer therapy (date of first subsequent anticancer systemic

therapy) and up to and including 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, NED (applies only to those patients entering the study with no disease at baseline), 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 8 weeks minus 1 week (to allow for an early assessment within the assessment window), after randomization (Parts 2A and 3) or first dose (Parts 1, 2B and 4). For CR/PR, the initial overall visit assessment that 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 site investigator data up until the first progression event, or first day of subsequent anticancer therapy, or last RECIST record in case of no PD or subsequent therapy. The denominator will be consistent with those used in the ORR analysis.

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 ≤17 weeks for Part 1-3 or ≤19 weeks for Part 4 (i.e. 16/18 weeks + 1 week to allow for a late assessment within the assessment window) after randomization (Parts 2A and 3) or first dose (Parts 1, 2B and 4), then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >17 weeks (Parts 1-3) or >19 weeks (Part 4) after randomization (Parts 2A and 3) or first dose (Parts 1, 2B and 4) then BoR will be assigned to the NE category.

A patient will be classified as a responder if the RECIST criteria for a confirmed CR or PR are satisfied at any time following randomization (Parts 2A and 3) or first dose (Parts 1, 2B and 4), prior to RECIST progression and prior to starting any subsequent anticancer systemic therapy.

Subjects who achieve CR or PR as determined by RECIST 1.1 will be included for the analysis of duration of response (DoR).

3.3.2.6 Blinded Independent Central Review (BICR)

The imaging scans will be reviewed by 2 primary radiologist reviewers using RECIST 1.1. If the overall timepoint assessments differ at any timepoint between the 2 primary reviewers, the case will be adjudicated by a third radiologist who must choose all the overall timepoint

assessments from the primary reviewer with which they more agree. If the overall timepoint assessments are identical between the 2 primary reviewers, the timepoint responses from the reviewer who completed their assessment of baseline scans first will be used for our analyses. For each patient, the BICR will define the overall visit response data (CR, PR, SD, PD, or NE) and the relevant scan dates for each time point (i.e., for visits where response or progression is/is not identified).

BICR data will only be analyzed for Part 2, Part 3 and Part 4 of the study.

3.3.2.7 RECIST 1.1-based efficacy endpoints

Analysis of the endpoints BoR, ORR, DCR, TTR, DoR, TTP, PFS will be based on Investigator and BICR assessments using RECIST 1.1.

All RECIST 1.1 assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues study drug(s) or receives another anticancer therapy.

At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD, or PD depending on the status of their disease compared with baseline and previous assessments. Baseline will be assessed before or on randomization date (Parts 2A and 3) or first dose date (Parts 1, 2B and 4). 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).

3.3.3 Efficacy Analyses

All efficacy endpoints will be summarized and analyzed using the Full Analysis Set. ORR will be additionally analyzed in the Response Evaluable Population.

Summaries and analyses will be performed for each study part [Part 1(1A+1B), Part 2 (2A+2B), Part 3, Part 4] separately as well as for Part 2 and Part 3 combined, and additionally for the China cohort (exceptions are: RECIST Investigator based analyses and BoR analyses will not be split by Part 2 and 3, but only presented for Part 2 and 3 combined). The efficacy data will be presented by randomized/allocated treatment arm for Parts 2, 3, 4 and China cohort, and by actual treatment arm for Part 1.

Efficacy subgroup analyses are defined in Section 3.3.4.

3.3.3.1 Best objective response

Best objective response (BoR) will be summarized with the number and percentage of subjects for the following categories: complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), and non-evaluable (NE). No statistical analyses are planned for BoR. BoR will be summarized in the FAS.

The following BoR analyses will be produced in FAS:

- BICR confirmed RECIST1.1 (Parts 2 and 3 combined, Part 4)
- BICR confirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined)
- Investigator confirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- BICR unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 4)
- Investigator unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)

3.3.3.2 Objective response rate

Only descriptive summaries of ORR including 95% CIs will be presented for each treatment arm. ORR will be summarized in the FAS and in the Response Evaluable Population. Confirmed and unconfirmed ORR will be analyzed.

The following ORR analyses will be produced in FAS and Response Evaluable Population:

- BICR confirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- BICR confirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined)
- [redacted]
- BICR unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- BICR unconfirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined, Part 2, Part 3)
- Investigator confirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- Investigator unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- BICR confirmed RECIST1.1 subgroup analysis (Parts 2 and 3 combined, Part2, Part 3, Part 4), FAS only

3.3.3.3 Disease control rate

Disease control rate (DCR) will be estimated with an exact 95% CI. DCR will be summarized in the FAS.

The following DCR analyses will be produced in FAS:

- BICR confirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- Investigator confirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 2, Part 3, Part 4, China cohort)

3.3.3.4 Time to response

The median TTR and its 95% CI will be assessed using the Kaplan-Meier method. TTR will be summarized and analyzed based on the FAS.

The following TTR analyses will be produced in FAS:

- BICR confirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- BICR confirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined, Part 2, Part 3)
- CCI [redacted]
- BICR unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- BICR unconfirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined, Part 2, Part 3)
- Investigator confirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- Investigator unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- BICR confirmed RECIST1.1 subgroup analysis (Parts 2 and 3 combined, Part 2, Part 3, Part 4)

3.3.3.5 Progression-free survival

For progression-free survival (PFS) based on PD due to radiographic PD, PD date will be the radiographic PD date.

The median PFS and its 95% CI will be estimated using the Kaplan-Meier method. PFS will be summarized and analyzed based on the FAS.

The following PFS analyses will be produced in FAS:

- BICR RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- CCI [redacted]

- Investigator RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 2, Part 3, Part 4, China cohort)
- BICR RECIST1.1 subgroup analysis (Parts 2 and 3 combined, Part2, Part 3, Part 4)

3.3.3.6 Duration of response

Descriptive data will be provided for the DoR in responding patients, including the associated Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached). DoR will be summarized and analyzed based on the FAS, for a subset of patients who achieved a confirmed response.

The following DoR analyses will be produced in FAS:

- BICR confirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- BICR confirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined, Part 2, Part 3)
- CCI
- BICR unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- BICR unconfirmed RECIST1.1 by line of therapy (Parts 2 and 3 combined, Part 2, Part 3)
- Investigator confirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- Investigator unconfirmed RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 4, China cohort)
- BICR confirmed RECIST1.1 subgroup analysis (Parts 2 and 3 combined, Part 2, Part 3, Part 4)

3.3.3.7 Time to progression

The median time to progression (TTP) and its 95% CI will be estimated using the Kaplan-Meier method. TTP will be summarized and analyzed based on the FAS.

The following TTP analyses will be produced in FAS:

- BICR RECIST1.1 (Parts 2 and 3 combined, Part 2, Part 3, Part 4)
- Investigator RECIST1.1 (Parts 2 and 3 combined, Part 1, Part 2, Part 3, Part 4, China cohort)

3.3.3.8 Overall survival

The median OS and its 95% CI will be estimated using the Kaplan-Meier method. The landmark 1-year OS (OS-12m), 18-month OS (OS-18m) and 24-month OS (OS-24m) will be analyzed in FAS using the method as outlined for analysis of OS. OS will be summarized and analyzed based on the FAS.

Additionally a separate summary will be produced for OS censored patients with total follow-up time less than or equal to median OS.

3.3.3.9 Change from Baseline in Tumor Sizes

The percent change from baseline in target tumor size will be calculated at each evaluable post-baseline disease assessment. It will be presented by subject using spider plots. The best percent change from baseline in target tumor is defined as the largest reduction or smallest increase (in the case where a reduction does not occur) from baseline observed over all post-baseline disease assessments and will be presented using waterfall plots.

CCI

CCI [redacted]

3.3.4 Subgroup Analyses

CCI [redacted]

- [redacted]
- [redacted]
- Sex (Male, Female)
- Age (<65, >=65 years)
- Line of therapy - treatment with sorafenib/VEGFR TKI (prior, no prior)
- CCI [redacted]
- Macrovascular invasion MVI (Yes, No)
- Extrahepatic spread EHS (Yes, No)
- MVI and EHS (MVI=Yes and/ or EHS=Yes, MVI=Yes and EHS=Yes, MVI=Yes and EHS=No, MVI=No and EHS=Yes, MVI=No and EHS=No, Missing either MVI or EHS)
- Region group: Asia (except Japan), Rest of the world (includes Japan).

All subgroups will be used for OS, PFS, ORR and DoR analyses in Parts 2 and 3.

PD-L1 subgroup will be used for Part 4 data analysis.

3.3.5 Exploratory Efficacy Analyses

CCI [redacted]

3.4 Patient Reported Outcomes

Not Applicable.

3.5 Pharmacodynamic Endpoints and Analyses

The pharmacodynamic analyses will be conducted by the MedImmune Translational Sciences group or designee. These analyses are outside the scope of the SAP.

3.6 Safety Analyses

Safety data will include AEs (including non-serious AEs and SAEs), discontinuation of investigational product due to toxicity, the actual value and change from baseline in laboratory test (including liver and viral labs), ECGs, vital signs, and ECOG performance status. Safety data will be summarized for each study part separately as well as for Part 2 and Part 3 combined.

"On treatment" will be defined as assessments between date of start dose and 90 days following the date of the last dose of study drug(s) (i.e., the last dose of durvalumab, tremelimumab or bevacizumab) or up to the date of initiation of the first subsequent anticancer systemic therapy, whichever occurs first. This definition applies to all safety reporting, unless otherwise specified.

AEs and SAEs will be collected from the time of signed informed consent through 90 days after the last dose of durvalumab, tremelimumab or bevacizumab. Any AEs that are unresolved at the subjects last AE assessment or other assessment/visits as appropriate in the study are followed up by the investigator for as long as medically indicated, but without further recording in the eCRF.

All AEs will be listed and treatment-emergent AEs (TEAEs) will be included for descriptive statistics. Treatment-emergent adverse events are defined as adverse events with onset or

worsening on or after first dose up to and including 90 days after last dose of study medication, or the initiation of subsequent anticancer therapy (date of first subsequent anticancer systemic therapy), whichever is earlier.

Laboratory test results, ECG, vital signs, and ECOG data will be summarized by scheduled time of evaluation and by treatment arm for each viral status cohort (HBV vs HCV vs uninfected) in Safety Analysis Set.

3.6.1 DLT Evaluation

The number and percentage of subjects with DLT will be presented based on the DLT Evaluable Set in Part 1A patients.

3.6.2 Adverse Events and Serious Adverse Events

Adverse events will be coded by MedDRA (using the latest or current MedDRA version) and graded according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE version 4.03). The type incidence, severity and relationship to study investigational product will be summarized. Specific adverse events will be counted once for each subject for calculating percentages. In addition, if the same adverse event occurs multiple times within a particular subject, the highest severity and level of relationship observed will be reported. If any associations of interest between adverse events and baseline characteristics are observed, additional stratified results may be presented. All treatment-emergent adverse events will be summarized overall, as well as categorized by MedDRA System Organ Class (SOC) and Preferred Term (PT).

The number and percentage of subjects reporting TEAEs will be reported by frequencies of preferred terms, MedDRA SOC, grade of adverse events, relatedness to the investigational product. Similar analysis will also be performed for serious adverse events. Adverse Events will be summarized in Safety Analysis Set.

Adverse events requiring systemic steroids will be summarized like other AEs, and they will be defined based on the information from the Concomitant Medications pages from eCRF.

Missing or incomplete start and end dates will be imputed as described in Section 3.1.

3.6.3 Adverse Events of Special and Possible Interest and Immune Mediated Adverse Events

Adverse events of special interest (AESI) and Adverse Events pf Possible Interest (AEPI) and immune-mediated Adverse Events (imAE) include AEs that have been described in study

Protocol Section 5.3 and some AEs that have been considered AESI or AEPI or imAE to the durvalumab program. They will be identified based on an external list of preferred terms created for the durvalumab program and they will be a subset of TEAEs in the study.

The AESI and AEPI and imAEs will be summarized similarly as other AEs described in Section 3.6.2.

Additionally the following categories of patients will be summarized based on the information from an external file from the Safety Specialists:

- Received systemic corticosteroids
- Received high dose steroids
- Received endocrine therapy
- Received other immunosuppressants.

3.6.4 Other significant Adverse Events

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs. 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 AEs (OAEs) and reported as such in the CSR. A similar review of laboratory, vital signs, and 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.

3.6.5 Deaths and Treatment Discontinuations due to Adverse Events

AEs that result in permanent discontinuation of investigational product or death will be summarized descriptively.

3.6.6 Clinical Laboratory Evaluation

Descriptive statistics will be provided for the clinical laboratory results and changes from baseline by scheduled time of evaluation including end of treatment visit as well as for the maximum and minimum post-baseline values.

Laboratory abnormalities will be graded according to the NCI CTCAE v4.03, if applicable. Frequencies of maximum observed grade will be presented for each laboratory parameter as well as the rates of subjects with Grade 3-4 toxicities. A shift table, presenting the 2-way

frequency tabulation for baseline and post-baseline grade at scheduled time of evaluation as well as the worst post-baseline grade, will be provided for clinical laboratory tests.

Shift tables for laboratory values by worst CTCAE 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 CTCAE grade shift outputs will be produced are:

- Haematology: Haemoglobin; Leukocytes; White blood cell count with differential [Lymphocytes (count, absolute); Neutrophils (count, absolute)]; Platelets; Fibrinogen; Prothrombin time/international normalization ratio; Activated partial thromboplastin time

- Clinical chemistry: ALT, AST, ALP, Total Bilirubin, Albumin, Magnesium – hypo and –hyper, Sodium – hypo and – hyper, Potassium – hypo and – hyper, Calcium – hypo and – hyper, Glucose – hypo and – hyper, Creatinine, Lipase, Amylase.

Liver Enzyme Elevations and Hy's law

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

a) ALT elevation: >= 3xULN - <= 5xULN, > 5xULN - <= 8xULN, > 8xULN - <= 10xULN, > 10xULN - <= 20xULN, >20xULN;

b) AST elevation: >= 3xULN - <= 5xULN, > 5xULN - <= 8xULN, > 8xULN - <= 10xULN, > 10xULN - <= 20xULN, >20xULN;

c) AST or ALT elevation: >= 3xULN - <= 5xULN, > 5xULN - <= 8xULN, > 8xULN - <= 10xULN, > 10xULN - <= 20xULN, >20xULN;

d) Total bilirubin elevation: >= 2xULN - <= 3xULN, > 3xULN - <= 5xULN, >5xULN;

e) ALT or AST and Total Bilirubin elevation:

   a. (ALT >= 3 x ULN or AST >= 3 x ULN) and (BILI $\geqslant$ 1.5 x ULN within 14 days on or after at elevation)

   b. (ALT >= 3x ULN or AST >= 3x ULN) and (BILI >= 2 x ULN within 14 days on or after at elevation)

   c. (ALT >= 3 x ULN or AST >= 3 x ULN) and ((BILI >= 2 x ULN and no ALP >= 2 x ULN) within 14 days on or after at elevation)

f) ALP elevation: >= 1.5xULN, > 3xULN.

Individual patient data with ALT or AST and Bilirubin elevations at any time will be listed.

Abnormal Thyroid function

Elevated TSH will be summarized per treatment group in terms of number (%) of patients with:

a) TSH > ULN

b) TSH > ULN with TSH <= ULN at baseline

c) TSH > 3xULN with TSH <=ULN at baseline

d) TSH > 10xULN

e) TSH > 10xULN with TSH <=ULN at baseline

f) TSH < Lower Limit of Normal (LLN)

g) TSH < LLN with TSH >= LLN at baseline

A shift table with thyroid tests change from baseline to maximum and from baseline to minimum value during treatment will be created.

3.6.7 Other Safety Evaluations

3.6.7.1 Vital Signs

Vital signs will be assessed at baseline and throughout the study. Descriptive statistics will be provided for the vital signs measurements and changes from baseline to scheduled time of evaluation including end of treatment visit as well as for the maximum and minimum post-baseline values.

3.6.7.2 Electrocardiogram

Electrocardiogram (ECG) parameters will be assessed at baseline as well as throughout the study. The ECG parameters (PR, RR, QRS, QT, and QTcF) will be summarized using descriptive statistics; changes from baseline to scheduled time of evaluation including end of treatment visit and to the maximum post-baseline values will be summarized.

For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point. If there are more than 3 assessments per time point (visit based on windows/day), the mean of all available values will be used.

For baseline calculation we only use assessments prior to the first dose of study medication, and calculate the mean of them if there is more than one assessment per time point.

QTcF (QT interval corrected for using Fridericia's formula) will be derived during creation of the reporting database using the reported ECG values (RR and QT) using the following formula: $QTcF = QT/RR^{(1/3)}$ where RR is in seconds

3.6.7.3 ECOG Performance Status

Eastern Cooperative Oncology Group (ECOG) performance status will be assessed at baseline as well as throughout the study. ECOG actual values will be summarized by scheduled time of evaluation and by treatment arm including end of treatment visit.

3.6.7.4 Child-Pugh Scores

Child-Pugh scores will be assessed at baseline as well as throughout the study.

- Child-Pugh score (scores 5 – 15): actual value over scheduled timepoints will be summarized using descriptive statistics.
- Child-Pugh class (A, B, C): classification over schedule timepoints will be summarized using descriptive statistics.

3.7 CCI

CCI

CCI

3.8 CCI

CCI

4 INTERIM ANALYSIS

Several interim analyses will be conducted for the study. The first interim analysis (IA1) is planned to occur when Part 1A and Part 1B are fully enrolled and subjects have been followed for at least 16 weeks. The purpose of the interim analysis following the completion of Part 1 enrollment is to determine whether the durvalumab and tremelimumab combination therapy has an appropriate benefit-risk ratio compared with either MEDI4736 or tremelimumab monotherapies and whether there is sufficient evidence to open Part 2 enrollment. As described in Protocol Section 3.2.4, the clinically "interesting" ORR is ≥ 20%. Therefore, if the probability of ORR ≥ 20% is < 20%, then enrollment in Part 2 will not proceed; if the probability of ORR ≥ 20% is > 80%, then Part 2 will open for enrollment. The probability is computed based on Bayesian approach. Under the rule, based on 36 subjects, including 12 subjects from Part 1A and 24 subjects from Part 1B, if ≤ 4 responses (confirmed PR or better) out of the 36 subjects

are observed, then enrollment in Part 2 will be terminated, and if ≥ 9 responses (confirmed PR or better) out of the 36 subjects are observed then enrollment will begin in Part 2. If the number of responses is 5, 6, 7 or 8 out of 36 subjects, then all available data (AEs, SAEs, duration of response, DCR, CCI etc.) will be evaluated by the sponsor and investigators to determine if the benefit-risk profile supports enrollment of Part 2.

The probabilities of different scenarios for different true response rates can be found in Table 4-1.

Table 4-1 Probabilities of Different Scenarios for Different True Response Rates

True Response Rate	Probability of Terminating Enrollment	Probability of Continuing Enrollment	Probability of Further Evaluation
0.05	97%	0%	3%
0.10	71%	1%	28%
0.15	36%	8%	56%
0.20	13%	28%	59%
0.25	3%	56%	40%
0.30	1%	80%	20%
0.35	0%	93%	7%

The second interim analysis (IA2) will occur after approximately 10 subjects per treatment arm in Part 3 have completed 4 weeks of follow-up. The objective of this interim analysis is to assess futility and safety of the monotherapy arms (Arms B and C) and safety of the additional durvalumab and tremelimumab combination therapy arm (Arm D). All available data from Part 2 and Part 3 will be included in the analysis. The goal of the second interim analysis is to assess the safety profiles of the four treatment arms. Furthermore, the preliminary efficacy of the monotherapy arms (Arms B and C) will be assessed to rule out futility. Futility will be indicated if there are no confirmed complete or partial responses observed in the treated subjects. This criterion was chosen because a true ORR of at least 10% is considered effective. This is greater than the 2% ORR seen on sorafenib (Llovet et al, 2008). If the ORR is 10% the probability of stopping a treatment arm with 36 subjects is 2%.

An additional ‘ad hoc’ analysis beyond the scope of the SAP was performed after IA2. The objective of this analysis was to access the safety and efficacy of the treatment arms to support the design and development of other future HCC studies for internal purposes. This analysis was essentially a re-analysis of IA2, but with additional follow-up data. For internal documentation purposes, this analysis is referenced as IA3.

The fourth interim analysis (IA4) was based on subjects included in the database as of the data-cutoff date of 28th Feb 2019. The objective of this analysis was to assess the safety and efficacy of the treatment arms to support the design and development of other future HCC studies for internal purposes. The subset for the analysis will include all patients from Parts 2 and 3 randomized (Parts 2A and 3) or enrolled (Part 2B) by 31st Dec 2018, except for one global Part 2A patient from China, who was excluded from analysis due to the complexity of dealing with additional vendors in China.

The fifth interim analysis (IA5) will be based on data cutoff data of 2nd Sep 2019. The objective of this analysis is to support the submission of a Phase 3 study HIMALAYA (D419CC000002). The scope of this analysis will be similar to IA1 for the HIMALAYA study. The subset for the analysis will include all patients from Parts 2 and 3 randomized (Parts 2A and 3) or enrolled (Part 2B) by 2nd Sep 2019, except for patients from the China tail.

The sixth interim analysis (IA6) will be based on data cutoff data of 28th Feb 2020. The objective of this analysis is to support the submission of a Phase 3 study HIMALAYA (D419CC000002). The scope of this analysis will be similar to IA2 for the HIMALAYA study. The subset for the analysis will include all patients from Parts 2A, 2B and 3 who were randomized/allocated to treatment on or prior to 2nd April 2019 except for patients from the China tail in Part 2A.

Additional interim analyses with the same aims may be performed to support the ongoing Phase 3 development for Parts 1-3. No formal adjustments will be made to the significance level used for assessment of multiple endpoints at different time-points.

Subject's tumor samples will be analyzed for PD-L1 expression and the prevalence of PD-L1 expression (low vs. high) will be evaluated as described in Section 3.3.3.10. Additional ad hoc analyses beyond the scope of the SAP may be performed to support the design and development of other future HCC studies for internal purposes. Such analyses will be documented outside of the SAP.

5 CHANGES OF ANALYSIS FROM PROTOCOL

- All efficacy analyses are based on the Full Analysis Set, instead of the Response Evaluable Population. Only the ORR analysis will be based on both FAS and Response Evaluable Population.

- "As-treated Population" was renamed to "Safety Analysis Set" to follow AZ and global standards.
- Overall survival, Progression Free Survival and Time to Progression definitions were modified to calculate time from randomization for randomized study parts (Parts 2A and 3), instead of from date of first dose of study drug for all patients.
- QTcB ECG parameter will not be summarized. The list of ECG parameters to summarize was created to match the list in HIMALAYA Phase 3 study.
- DCR-24w endpoint will not be included in CSR and all references to it were removed from SAP. It was initially added to CSP to support the Phase 3 HIMALAYA study. For publications perspective only an overall DCR is of interest.
- Landmark PFS-6m and TTP-6m will not be analyzed as it is not informative in this study.

6 REFERENCES

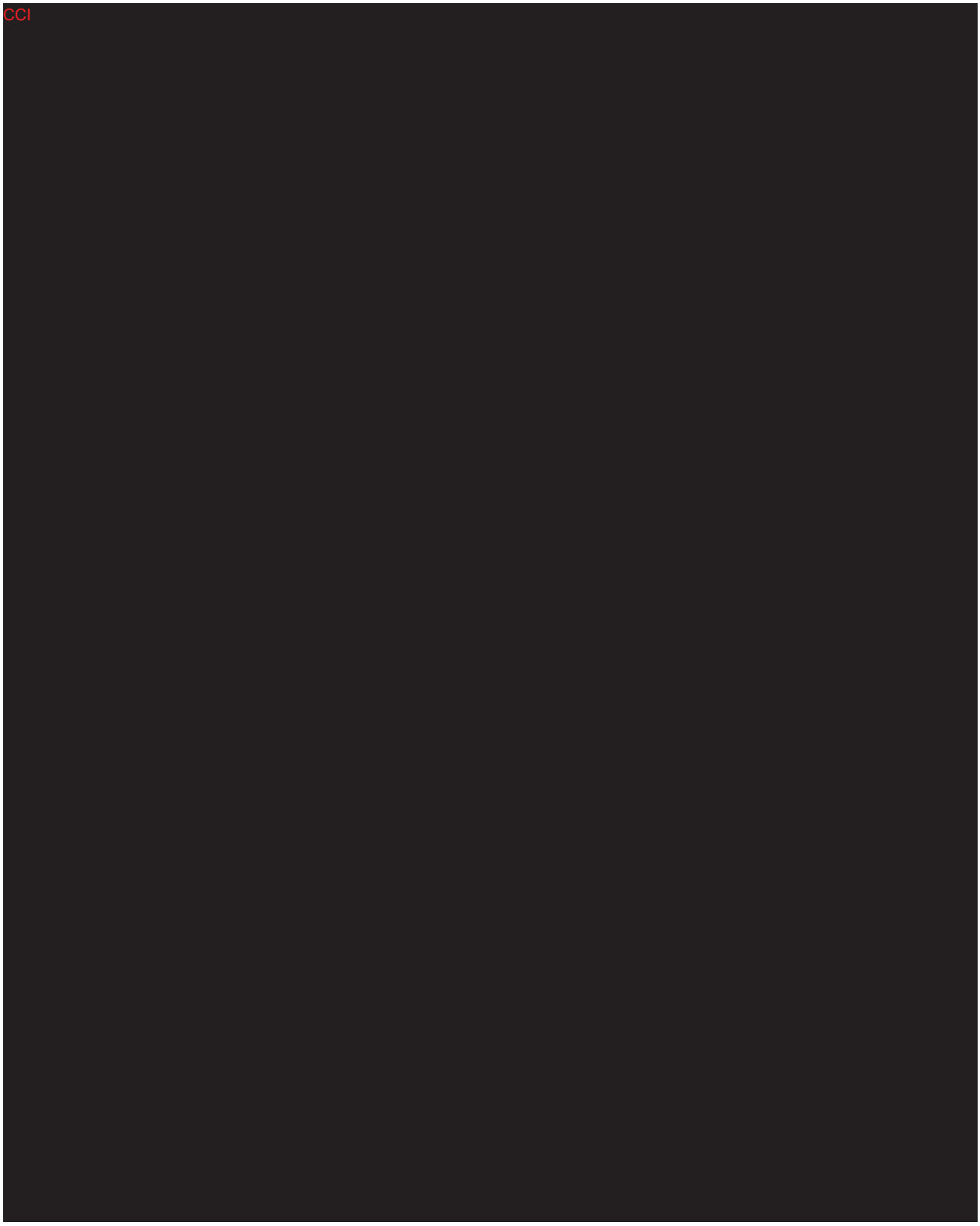
Common Terminology for Adverse Events, Version 4.03. United States Department of Health & Human Services, National Institutes of Health, National Cancer Institute, Washington, DC, 2009.
http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf. Downloaded May 10, 2012.

Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. United States Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). Washington, DC. May, 2007.
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf. Downloaded August 9, 2012.

CCI [redacted]

Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008 Jul 24;359(4):378-90.

7 APPENDIX




CCI

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d4190c00022-sap-ed-8		
Document Title:	Statistical Analysis Plan Edition 8	
Document ID:	Doc ID-004007104	
Version Label:	6.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
07-Dec-2020 15:17 UTC	PPD	Content Approval
07-Dec-2020 12:04 UTC		Author Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.