Statistical Analysis Plan	
Study Code	D5082C00003
Edition Number	2.0
Date	28th August 2019

A Phase III, Open Label, Randomised, Controlled, Multi-Centre, Study to Assess the Efficacy and Safety of Savolitinib versus Sunitinib in Patients with MET-Driven, Unresectable/Locally Advanced or Metastatic Papillary Renal Cell Carcinoma (PRCC)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

The following is a list of abbreviations used in this document. Each individual SAP would contain its own set of abbreviations.

Abbreviation or special term	Explanation
AE	Adverse event
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
A&R	Analysis and reporting
BICR	Blinded independent central review
BL	Baseline
BoR	Best objective response
BP	Blood pressure
CI	Confidence interval
CR	Complete response
CRF	Case report form
CRO	Contract research organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CTCAE-PRO	Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events
[redacted]	[redacted]
CV	Coefficient of Variation
DAE	Discontinuation of investigational product due to adverse events
DBL	Database lock
DCO	Data cut-off
DCR	Disease Control Rate
DNA	Deoxyribonucleic acid
DoR	Duration of response

Abbreviation or special term	**Explanation**
d.p.	Decimal place
ECG	Electrocardiogram
ECHO	Echocardiogram
eCRF	Electronic case report form
EDoR	Expected duration of response
EOT	End of Treatment
ePRO	Electronic Patient Reported Outcomes
EQ-5D-5L	EuroQoL five dimensions, five level
EQ-VAS	EuroQoL-Visual Analogue Scale
EuroQoL	European Quality of Life
FACT-G	Functional Assessment of Cancer Therapy – General
FACIT-F	Functional Assessment of Chronic Illness Therapy-Fatigue
FAS	Full Analysis Set
FDA	Food and Drug Administration
FFPE	Formalin Fixed and Paraffin Embedded
FH	Fumarate Hydratase
FKSI-19	Cancer Therapy Kidney Symptom Index-19
HGF	Hepatocyte growth factor
HR	Hazard ratio
HRQoL	Health related quality of life
ICF	Informed Consent Form
IDMC	Independent Data Monitoring Committee
IMDC	International Metastatic RCC Database Consortium
IP	Investigational Product
ITT	Intention-to-treat
IVRS	Interactive Voice Response System
LD	Longest diameter
LFT	Liver Function tests
LRCI	Likelihood ratio confidence interval
LVEF	Left Ventricular Ejection Fraction
MedRA	Medical Dictionary for Regulatory Activities

Abbreviation or special term	**Explanation**
MES	Molecular Epidemiology Study
MET	Met proto-oncogene (hepatocyte growth factor receptor)
MRI	Magnetic resonance imaging
MTP	Multiple-testing procedure
MUGA	Multigated acquisition scan
NA	Not applicable
NCI	National Cancer Institute
NE	Not evaluable
NED	No evidence of disease
NGS	Next Generation Sequencing
NTL	Non-target lesion
OAE	Other significant adverse events
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PID	Percentage intended dose
PFS	Progression free survival
PFS2	Second progression-free survival
PK	Pharmacokinetics
PO	Orally
PR	Partial response
PRCC	Papillary renal cell carcinoma
PRO	Patient reported outcome
PT	Preferred Term
QD	Once a day
QOL	Quality of life
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumours
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Safety Analysis Set

Abbreviation or special term	**Explanation**
SAS®	A commercially available integrated system of software products, commonly used for reporting and analysis of Clinical Studies
SD	Stable disease
SOC	System Organ Class level of MedRA
TBL	Total Bilirubin
TL	Target lesion
tsm	Tumour size measurements
ULN	Upper limit of normal
VAS	Visual analogue scale
VEGF TKI	Vascular endothelial growth factor receptor tyrosine kinase inhibitor
VHL	Von Hippel‑Lindau
TNM	Tumour, Node and Metastasis
AJCC	American Joint Committee on Cancer
WHO	World Health Organisation
ATC	Anatomical Therapeutic Chemical
MedDRA	Medical Dictionary for Regulatory Activities
SOC	System Organ Class
PT	Preferred Term
CTA	Clinical Trial NGS assay
FMI	Foundation Medicine Inc

AMENDMENT HISTORY

Date	**Brief description of change**
21st June 2018	Version 1.1 Updated by Phastar prior to IA for MES and prior to BDR1. Summary of changes: • Study flow chart updated in line with CSP V3.0 • MMRM added for EQ5D index/VAS • New sections added describing patient disposition, demographic, concomitant medication, ECG ECHO/MUGA, handling of partial dates and EQ5D • Additional analysis set for all patients added to section 2 • Updated text for when we have multiple values on the same day for visit based summaries to take the average/worst result depending on value • New subgroup added: MET-driven criteria • New subgroup analysis added: ORR descriptive statistics • Step 3 updated for TL visit responses subsequent to CR section • Treatment compliance definition added • Analysis methods section updated in line with updated shells • Modified total safety follow up to include 'date of first dose of subsequent anti-cancer therapy' • Sensitivity analysis of DoR updated in line with CSP
12th February 2019	Version 1.2 Updated by Phastar to incorporate changes to the clinical study protocol (v6, dated 19 December 2018) following from the decision to terminate study recruitment. See details below. Summary of changes: • Minor formatting changes (such as re-numbering to incorporate new sections) have been applied thoroughout. • Section 1.2 (Study Design) – text has been added in relation to the study termination • Section 1.3 (Number of subjects) – Rationale added as to why recruitment into the study has been terminated and

	why the interim analysis and sample size-reassessment will not be conducted. • New section added, Section 1.5 (Patient management post final analysis) • Section 3.2.2 (Overall survival) – text amended to state that survival calls will be within 5 days to match the protocol • Section 3.2.9.4 (PRO CTCAE) – text added to clarify which countries will receive the questionnaire • Section 4.2 (Analysis Methods) – Table 5 has been removed as it is no longer applicable. Additionally the sections on multiplicity and the proportionality assumption have been removed. • Section 4.2.2.1 (PFS sensitivity analyses) – References to evaluation bias and attrition bias have been removed. The additional support summaries have also been cut down to match the amended protocol text. • Section 4.2.3 (Overall survival) – text amended to state that survival calls will be within 5 days to match the protocol. The sensitivity analyses has been removed. • Sections 4.2.4 – 4.2.8, 4.2.10 have been revised to clarify the analysis of ORR, DoR, PFS2, Change in TL tumour size, DCR and PRO's. • Section 4.2.9 (Subgroup analysis) – this has been amended to state there will be no subgroup analyses. • Section 4.2.14. (Adverse events) – summaries of AE's leading to hospitalisation, dose interruption, dose reduction and OAE's have been removed. • Section 4.2.14.3 (AESI) – this section has been revised to clarify the analyses that will be performed. • Section 5 (Interim analyses) – updated to explain interim analyses will no longer take place with the exception of interim reviews of safety data by IDMC, which will continue until the end of the study. • New section added, Section 6 – Changes from protocol. Updated following AZ review of the amended SAP text (15th March 2019). Summary of changes: • Minor formatting changes resulting from the review (such as re-numbering of sections, updating of tense used and

	repositioning or re-wording of text) have been applied throughout. • The Other significant AE's (OAEs) sub-header within section 3.3.5 has been removed as an OAE output table is no longer required. • Section 4.2.6 (Time from Randomisation to second progression or death (PFS2)) and Section 6 (Changes of Analysis from Protocol) have been updated to record the removal of PFS2 analyses. • Section 4.2.10 (Patient reported outcomes (PROs)) has been updated to state that PRO-CTCAE will be listed over time (in earlier versions it stated summarised descriptively). • Reference to the production of graphical plots for the PRO endpoints in Section 4.2.10 (Patient reported outcomes (PROs)) and Section 6 (Changes of Analysis from Protocol) has been removed. SAP text moved into the updated AZ template.
28th August 2019	Version 2.0 No changes to content made – updated version number to reflect final status.

1 STUDY DETAILS

1.1 Study Objectives

1.1.1 Primary Objective

Primary Objective:	**Outcome Measure:**
To determine the efficacy of savolitinib when compared to sunitinib in patients with MET driven, unresectable and locally advanced, or metastatic PRCC	PFS (time to earliest progression as defined by RECIST 1.1 and confirmed by BICR, or death)

1.1.2 Secondary Objective

Secondary Objective:	**Outcome Measure:**
To compare the efficacy of savolitinib versus sunitinib in patients with MET-driven, unresectable and locally advanced, or metastatic papillary renal cell carcinoma (PRCC)	• Overall Survival (OS) • Objective Response Rate (ORR), duration of response (DoR) and best percentage change in tumour size by BICR using RECIST 1.1 criteria • Disease Control Rate (DCR) at 6 and 12 months
To assess the impact of savolitinib and sunitinib on disease related symptoms and health-related QOL in this patient population	• Mean change from baseline in FKSI-19 and FACIT-F scores
To evaluate the pharmacokinetics of savolitinib in this patient population	• PK concentration data

1.1.3 Safety Objective

Safety Objective:	**Outcome Measure:**
To evaluate the safety and tolerability of savolitinib in relation to sunitinib	• AEs/SAEs adverse events [AEs] as characterized and graded by National Cancer Institute [NCI] Common Terminology Criteria for Adverse Event [CTCAE] v4.03 • Collection of clinical chemistry/haematology parameters, liver function tests, echocardiograms and electrocardiograms (ECGs), vital signs including blood pressure (BP) and heart rate

1.1.4 Exploratory Objective

Exploratory Objective:	**Outcome Measure:**
To compare the efficacy of savolitinib versus sunitinib in patients with MET-driven, unresectable and locally advanced, or metastatic papillary renal cell carcinoma (PRCC)	Time from randomisation to objective disease progression on subsequent anti-cancer therapy after PD by RECIST 1.1 on study medication, or death (PFS2)
To assess the impact of savolitinib vs. sunitinib on patient reported AEs	Collection of PRO CTCAE symptoms
To investigate the health economic impact of treatment and the disease on hospital related resource use and health state utility	Number, type and reason for hospitalisations and unscheduled clinic visits, procedures undertaken and length of hospital inpatient stays Health state response and utility index derived from the EQ-5D-5L
[redacted]	[redacted]
To collect and store DNA (according to each country's local and ethical procedures) for future exploratory research into genes/genetic variation that may influence response (i.e., distribution, safety, tolerability and efficacy) to study medications and or susceptibility to disease (optional)	• Pharmacogenetic analyses on blood samples
[redacted]	[redacted]

The following exploratory outcome measures will not be reported in the clinical study report:

[redacted] Due to this no detailed analysis methods are provided in this document.

1.2 Study Design

This is an open-label, randomised, controlled, multicentre, phase III study designed to evaluate the efficacy of savolitinib compared to sunitinib in patients with MET-driven, unresectable and locally advanced, or metastatic PRCC. Patients can be treatment naïve or previously treated but cannot have previously received sunitinib or a MET inhibitor. Patients who have received prior systemic therapy must have had disease progression in soft tissue disease or bone within 6 months of the last dose of the most recent systemic therapy. Bone progression is defined by the appearance of ≥2 new lesions on bone scan. Supportive care agents such as denosumab or radium 223 are not considered prior systemic therapy, and patients who have received these agents are potentially eligible.

Once the patient has undergone the Part 1 screening visit and signed the Part 1 ICF, a formalin fixed and paraffin embedded (FFPE) tumour sample meeting the requirements specified in the Laboratory Manual will be sent to the sponsor-designated central laboratory. For first-line patients, the diagnostic specimen can be used. For previously treated patients, the most recent archival specimen obtained during a clinical procedure is preferable to the diagnostic specimen only if it meets specifications and has an equal or higher tumour cellularity. If the above is not available or unlikely to yield sufficient material for testing, the patient will have the option to provide an FFPE tumour tissue block from a de novo core needle tumour biopsy. The biopsied tumour lesion must not be a RECIST target lesion, as the biopsy may affect the accuracy of the imaging.

Locally available pathology results confirming PRCC will be used for study entry. All patients must have a confirmed MET-driven tumour by the sponsor designated central laboratory using Next Generation Sequencing (NGS) to evaluate for qualifying alterations prior to study entry.

A MET-driven tumour is defined as any of the following molecular alterations or combination of these alterations detected in tumour tissue using the validated MET NGS test in the absence of co-occurring FH or VHL mutations:

- Chromosome 7 gain
- MET amplification
- MET kinase domain mutations
- HGF amplification

The main study (Part 2) screening assessments must take place within 28 days of randomisation.

Patients who fulfil all the eligibility criteria will be randomised in a ratio of 1:1 to receive:

- Savolitinib 600 mg (400 mg if <50 kg) by mouth (PO) with a meal once daily (QD) versus
- Sunitinib 50 mg PO QD, with or w/o food.

For the purposes of planning, a 6-week period will be called a 'cycle'. Treatment with sunitinib will be given 4 weeks on/2 weeks off. Treatment with savolitinib will be given continuously.

Patients will be stratified based on the IMDC risk group criteria (Kroeger et al 2013), using the number of pre-defined risk factors to assign patients into favourable, intermediate or poor prognostic groups as well as whether they are treatment naive or previously treated with or without a VEGF TKI.

Following randomisation, patients will attend scheduled study visits as outlined in the study plan (Table 1 in the Study Protocol). Efficacy will be assessed by tumour assessments every 6 weeks, corresponding to the start of each cycle, and then every 12 weeks after the first year, until disease progression as defined by RECIST 1.1 and confirmed by BICR. All scans will be read by BICR after notification of Progressive Disease by the investigator. If PD is not centrally confirmed, each subsequent scan will be read by BICR once it is received and processed by the sponsor designated vendor. Depending on the speed and completeness of image submission to the sponsor-designated vendor, the turn- around time generally should be within 10 days.

During verification of the institution's diagnosis of radiographic progression by BICR participants should continue to receive study treatment if not medically contra-indicated. The decision whether to continue or to withhold study treatment will be at the discretion of the investigator, but if not medically contra-indicated, patients should be strongly encouraged not to start subsequent anti-cancer therapy until BICR-confirmation of PD. If disease progression is not confirmed by BICR, the patient may continue/resume study treatment and must continue to undergo assessments as per the study schedule, including tumour evaluations at 6-week intervals (or sooner if felt to be medically indicated) until disease progression is confirmed by BICR. Those patients who begin subsequent non-study anti-cancer therapy prior to BICR confirmed PD should continue to undergo tumour assessments, and sites must send each scan to the sponsor-designated vendor until PD is confirmed by BICR.

Regardless of initial treatment, patients will have two options for post progression therapy following BICR-confirmation of PD:

1. Receive subsequent non-study anti-cancer therapy that does not contravene local practice.
2. Continue to receive the assigned study treatment as long as in the opinion of the investigator, the patient is deriving benefit and the patient meets the original eligibility criteria in terms of Performance Status and laboratory values. This decision must be discussed with and approved by the AZ study physician. Such patients will continue to undergo monitoring as per the Study Plan (Table 1 of the study protocol).

There will be no cross-over.

An EOT visit will be conducted as soon as possible and preferably within 7 days following the decision to discontinue study medication.

The purpose of this visit is to:

a) Discuss any further treatment options and/or follow-up
b) Discuss the possibility of a biopsy, which should be strongly encouraged, particularly for patients who had responded to treatment (patients may have signed a consent line for this on the Main Study ICF)
c) Obtain the required EOT tests and collect the ePRO device after the last set of questionnaires are completed
d) Collect any unused study medication and study-medication containers from the patient

Once a patient permanently discontinues study drug, a 30-day (+/- 7 days) safety follow-up will be performed to follow-up on any SAE/AEs and concomitant medications (including any subsequent cancer therapy). If there are no ongoing SAE/AEs at the time of drug discontinuation, the safety follow-up can be done via telephone contact.

Determination of PD on subsequent anti-cancer therapy in patients who had PD by RECIST 1.1 on study medication (second progression-free survival, PFS2) will be by institutional call. Patients who choose to continue on their initial therapy after BICR-confirmed PD will not be considered to have a second PFS event until they discontinue study medication and progress on subsequent anti-cancer treatment.

In the overall survival (OS) follow-up period following PFS2, subsequent therapies and vital status will be documented at least every 12 weeks until death, lost to follow-up, withdrawal of consent, or end of study, whichever comes first.

A recent review of the final results of the Molecular Epidemiology Study (MES) concluded that in this data-set, [redacted]

Following a review of the MES results, a decision, endorsed by the Independent Data Monitoring Committee (IDMC), was made to terminate recruitment into SAVOIR on 22nd November 2018.

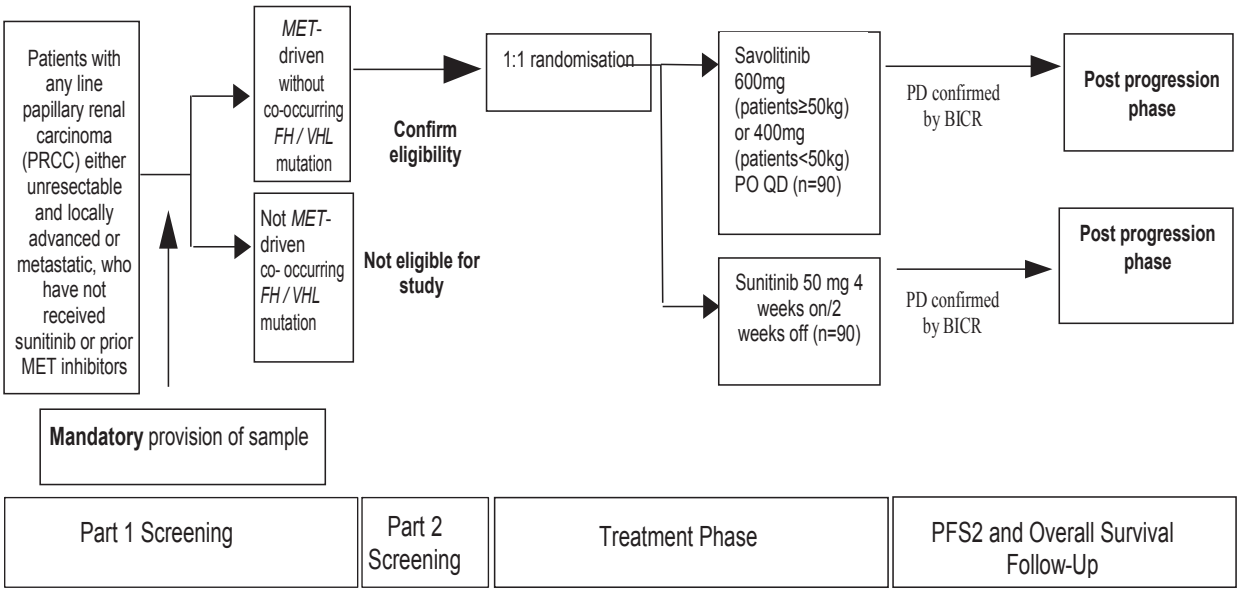
Patients randomised in SAVOIR may continue to receive their assigned treatment for as long as they are deriving clinical benefit or until they meet a protocol-defined stopping criterion and must follow the study plan as outlined. Patients already in screening (who have signed the ICF) will have the option to continue screening and to be randomised if eligible. In light of the MES results, patients currently on study must be re-consented after approval of ICF Addendum No. 1 dated 03 Dec 2018, as an update to Master ICF v5 dated 24 Jul 2018.

The final analysis of all endpoints will be conducted at the earliest time when the following two criteria are met:

1 36 PFS events by investigator assessment
2 The opportunity to have at least 7.5 months follow-up from randomisation

If criterion #1 has not been met 12 months after last subject in, the final analysis will be performed. At this time point, the clinical study database will close to new data.

Figure 1 Study Flowchart



BICR, Blinded Independent Central Review; FH/VHL, fumarate Hydratase/von Hippel-Lindau; PD, progression of disease; PFS2, second progression-free survival; PO, orally; PRCC, papillary cell renal carcinoma; QD, once a day.

1.3 Number of Patients

The initial assumption was that the true treatment effect had a hazard ratio (HR) of 0.6 (which would have translated to an approximate 3 month improvement in median PFS over an assumed 4 month median PFS for MET-driven patients on sunitinib, assuming PFS was exponentially distributed), 121 PFS events confirmed by BICR would have had to have been observed for the study to have 80% power to show a statistically significant difference in PFS at the two-sided 5% level. The smallest treatment difference that would have been statistically significant at the final analysis was a PFS HR of 0.69. Assuming 67% maturity; 180 patients would have been needed to be randomised.

An interim futility analysis of PFS was planned to be conducted when a total of approximately 36 BICR-confirmed PFS events had been observed (30% of 121 PFS events, which was

estimated to occur 17 months after the first patient had been randomised). Following IDMC endorsement to terminate recruitment into SAVOIR, the interim analysis will not occur. However, interim reviews of safety data by the IDMC will continue until end of study.

Recruitment to the study was closed on 22 November 2018. The randomisation was closed on 8 January 2019 and up to this date, 60 patients were randomised.

[redacted]

[redacted]

Following IDMC endorsement to terminate recruitment into SAVOIR, the sample size reassessment was not performed due to the [redacted]

1.4 Trial Integrity

This is an open-label study. However, steps will be taken to maintain data integrity in a well-controlled manner that is free of statistical bias, while also minimizing operational bias. A document was generated prior to the first patient randomised pre-specifying nominated individuals who will be granted access to any treatment-revealing data, with their purpose including any expectation of sharing outputs beyond those named. Further details are available in the Trial Integrity Document.

1.5 Patient Management Post Final Analysis

Patients receiving study drug, either savolitinib or sunitinib, will be allowed to remain on treatment for as long as they are deriving clinical benefit, in the opinion of investigator(s) or until meeting any discontinuation criteria.

Patients will be monitored in accordance with the investigator's standard clinical practice and national product label (for patients being treated with sunitinib). Dispensing of study treatment post-final analysis will be done outside of IVRS/IWRS. At routine clinic visits, patients will return partially used and unused medication, and a thorough drug accountability assessment will be performed at the site.

AstraZeneca will collect information (during the treatment period and for 30 (±7) days after discontinuation of study medication) on SAEs, overdose and pregnancy. Drug accountability information will be recorded in the source documents.

If an investigator learns of any SAEs, including death, at any time after a patient has completed the study, and he/she considers there is a reasonable possibility that the event is causally related to the investigational product, the investigator should notify AstraZeneca Patient Safety. Additionally, any SAE or non-serious AE that is ongoing at the time of this data cut-off, must be followed up to resolution unless the event is considered by the investigator to be unlikely to resolve, or the patient is lost to follow-up.

2 ANALYSIS SETS

2.1 Definition of Analysis Sets

2.1.1 All patients analysis set

This analysis set comprises all patients screened for the study and will be used for reporting of disposition.

2.1.2 Full Analysis Set (FAS)

The FAS will include all randomised patients with treatment groups assigned in accordance with the randomisation, regardless of the treatment actually received. Patients who were randomised but did not subsequently receive treatment are included in the FAS. The analysis of data using the FAS therefore follows the principles of intention-to-treat (ITT).

2.1.3 Safety analysis set (SAS)

The safety analysis set consists of all patients who received at least one dose of randomised treatment (regardless of whether that was the randomised therapy intended or indeed whether, in rare cases, they received therapy without being randomised), according to the treatment they actually received.

2.1.4 PK analysis set

All patients who receive at least 1 dose of savolitinib per the protocol, for whom any post-dose PK data are available and do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the PK analysis set. The population will be defined by the Study Team Physician, Pharmacokineticist and Statistician prior to any analyses being performed. Where a protocol deviation impacts only part of a patient's data, the affected portion of the patient's PK data will be excluded from PK analysis and summary statistics, and the remaining valid data will be utilized.

Table 1 Summary of outcome variables and analysis populations

Outcome variable	***Populations***
Efficacy data	
PFS	*FAS*
Efficacy Data	
OS, ORR*, duration of response*, best percentage change in tumour size, DCR, symptom/HRQoL endpoints	*FAS*
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	*FAS*
Baseline and disease characteristics	*FAS*
Important deviations	*FAS*
Medical/surgical history	*FAS*
Previous anti-cancer therapy	*FAS*
Concomitant medications/procedures	*FAS*
Subsequent anti-cancer therapy	*FAS*
PK Data	
PK data	*PK*
Safety data	
Exposure	*Safety*
Adverse events	*Safety*
Laboratory measurements	*Safety*
Vital signs	*Safety*
ECGs	*Safety*

**Patients who are evaluable for the analysis of ORR are those with measurable disease at baseline. Patients who are evaluable for the analysis of DoR are those who responded in the ORR analysis.*

2.2 Protocol Deviations

The following general categories will be considered important protocol deviations and will be programmatically derived from the eCRF data. These will be listed and discussed in the CSR as appropriate:

- Patients randomised but who did not receive study treatment (Deviation 1).
- Patients who deviate from key entry criteria (inclusion criteria 3, 4, 5, 7, and exclusion criteria 5, 14 and 21) per the Clinical Study Protocol (CSP) (Deviation 2).
- Baseline RECIST scan > 28 days before start date of randomised treatment (Deviation 3).
- No baseline RECIST 1.1 assessment on or before date of randomisation (Deviation 4).
- Received prohibited concomitant anti-cancer therapies (Deviation 5). Please refer to the CSP appendix H for those medications 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.
- Patients randomised who received their randomised study treatment at an incorrect dose or received an alternative study treatment to that which they were randomised (Deviation 6).

Patients who receive the wrong treatment at any time will be included in the safety analysis set as described in section 2.1. During the study, decisions on how to handle errors in treatment dispensing (with regard 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 and/or statistician.

The important protocol deviations will be listed and summarised by randomised treatment group. 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 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). A per-protocol analysis excluding patients with specific important protocol deviations is not planned; however, a 'deviation bias' sensitivity analysis may be performed on the progression free survival endpoint excluding patients with deviations that may affect the efficacy of the trial therapy if > 10% of patients in either treatment group:

- Did not have the intended disease or indication or
- Did not receive any randomised therapy.

The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock and will be documented prior to the primary analysis being conducted.

In addition to the programmatic determination of the deviations above, other study deviations captured from the CRF module for inclusion/exclusion criteria will be tabulated and listed. Any other deviations from monitoring notes or reports will be reported in an appendix to the CSR.

3 PRIMARY AND SECONDARY VARIABLES

3.1 Derivation of RECIST Visit Responses

For all patients, the RECIST tumour 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.

Baseline radiological tumour assessments are to be performed no more than 28 days before the start of randomised treatment and ideally as close as possible to the start of study treatment. Tumour assessments are then performed every 6 weeks (12 weeks after the first year) following the start of study treatment until disease progression.

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

A Blinded Independent Central Review (BICR) of all radiological imaging data will be carried out using RECIST version 1.1 and the overall visit response data (CR, PR, SD, PD or NE) will be provided at each visit (ie the reviewers will provide the overall visit response according to RECIST 1.1 and no programmatic derivation of visit response is necessary). Central confirmation of each scan will also be conducted.

From the investigator's review of the imaging scans, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD or PD, using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions and depending on the status of their disease compared with baseline and previous assessments. If a patient has had a tumour 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).

Please refer to section 3.1.1 for the definitions of CR, PR, SD and PD.

RECIST outcomes (i.e. PFS, ORR etc.) will be calculated programmatically for the site investigator data (see section 3.2) from the programmatically derived overall visit responses.

RECIST outcomes based on BICR data (i.e. PFS, ORR etc) will be calculated programmatically from the central review provided overall visits responses.

3.1.1 Target lesions (TLs) – site investigator data

Measurable disease is defined as having at least one measurable lesion, not previously irradiated, which is ≥ 10 mm in the longest diameter (LD), (except lymph nodes which must have short axis ≥ 15 mm) with computed tomography (CT) or magnetic resonance imaging (MRI) which is suitable for accurate repeated measurements. A patient can have a maximum of five measurable lesions recorded at baseline with a maximum of two lesions per organ (representative of all lesions involved and suitable for accurate repeated measurement) and these are referred to as target lesions (TLs). If more than one baseline scan is recorded then measurements from the one that is closest and prior to randomisation will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement. In which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

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

Note: For patients who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit responses will be based on the overall NTL assessment and the absence/presence of new lesions (see section 3.1.3 for further details). If a patient does not have measurable disease at baseline then the TL visit response will be not applicable (NA).

Table 2 TL Visit Responses (RECIST 1.1)

Visit Responses	Description
Complete response (CR)	Disappearance of all TLs. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10mm.
Partial response (PR)	At least a 30% decrease in the sum of diameters of TLs, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive disease (PD)	A ≥ 20% increase in the sum of diameters of TLs and an absolute increase of ≥ 5mm, taking as reference the smallest sum of diameters since treatment started including the baseline sum of diameters.
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 TLs 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 TL response.

Visit Responses	Description
Not applicable (NA)	No TLs are recorded at baseline.

Rounding of TL data

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

Missing TL data

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

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

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

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

Lymph nodes

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

TL visit responses subsequent to CR

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

- Step 1: If all lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node LD increases by 20% but remains < 10mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0mm or < 10mm for lymph nodes) then response will be set to NE

irrespective of whether, when referencing the sum of TL diameters, the criteria for PD are also met.

- Step 3: If not all lesions meet the CR criteria (i.e. a pathological lymph node selected as TL has short axis >10mm or the reappearance of previously disappeared lesion) or a new lesion appears then response will be set to PD
- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR

TL too big to measure

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

TL too small to measure

If a TL becomes too small to measure then this will be indicated as such on the case report form and a value of 5mm will be entered into the database and used in TL calculations. However a smaller value may be used if the radiologist has not indicated 'too small to measure' on the case report form and has entered a smaller value that can be reliably measured. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention

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

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolisation), 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 tumours:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD, this will remain as a valid response category.
- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and if ≤ 1/3 of the TLs have missing measurements then scale up as described in the 'Scaling' section below. If the

scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If, after both steps, PD has not been assigned, then, if appropriate (i.e. if ≤ 1/3 of the TLs have missing measurements), the scaled sum of diameters calculated in step 2 should be used, 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 have a value of 0 (or <10mm for lymph nodes) recorded. If scaling up is not appropriate due to too few non-missing measurements then the visit response will be set as NE.

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

Scaling (applicable only for irradiated lesions/lesion intervention)

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

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

Example of scaling

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

Lesion 5 is missing at the follow-up visit; it had a nadir measure of 29.3cm.
The sum of lesions 1-4 at the follow-up is 26 cm. The sum of the corresponding lesions at the nadir visit is 26.8 cm.

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

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

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

Lesions that split in two

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

Lesions that merge

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

Change in method of assessment of TLs

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

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

3.1.2 Non-target lesions (NTLs) and new lesions

At each visit, 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 overall assessment of NTLs as follows:

Table 3 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.
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 tumour burden has increased sufficiently to merit a determination of disease progression. A modest ‘increase’ in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

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

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

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

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.
If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present, but should not overtly affect the derivation.

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

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

3.1.3 Overall visit response

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

Table 4 Overall visit responses

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

3.1.4 Independent review

An independent review of all scans used in the assessment of tumours using RECIST 1.1 will be conducted. All radiological scans for all patients (including those at unscheduled visits, or outside visit windows) will be collected on an ongoing basis and sent to an external imaging

vendor for central analysis following completion of each scan. All scans will be read by BICR after notification of Progressive Disease by the investigator, death, or withdrawal from the study. If PD is not centrally confirmed, each subsequent scan will be read by BICR once it is received and processed by the imaging vendor, the result will then be sent to the A&R MARS provider. Patients without PD by investigator, BICR will be continued up to study withdrawal /DCO to ensure all patients have complete BICR assessments in the primary analysis. Prior radiotherapy and location will also be provided to the BICR to allow the selection of appropriate TLs.

The imaging scans will be reviewed by two independent radiologists using RECIST 1.1 and will be adjudicated, if required (i.e. two reviewers' review the scans and adjudication is performed by a separate reviewer in case of a disagreement). For each patient, the BICR will define the overall visit response (i.e. the response obtained overall at each visit by assessing TLs, NTLs and new lesions) data and no programmatic derivation of visit response is necessary. (For patients with TLs at baseline: CR, PR, SD, PD, NE; for patients with NTLs only: CR, SD, PD, NE; for patients with no disease identified at baseline: PD, no evidence of disease [NED], NE). If a patient has had a tumour assessment that cannot be evaluated then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the reviewer that the adjudicator agreed with will be selected as a single reviewer (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to report all BICR RECIST information including dates of progression, visit response, censoring and changes in target lesion dimensions. Endpoints (of ORR, PFS and DoR) will be derived programmatically from this information.

During verification of the institution's diagnosis of radiographic progression by BICR, participants may continue to receive study treatment. The decision whether to continue or to withhold study treatment will be at the discretion of the investigator, but if not medically contra-indicated, patients should be encouraged to not start alternative anti-cancer therapy until confirmation of PD by BICR. If PD is not confirmed by BICR, the patient may continue/resume study treatment and must continue to undergo assessments as per Table 1 of the CSP, including scans at 6 week intervals (or sooner if felt to be medically indicated) until PD is confirmed by BICR. Those who begin subsequent non-study anti-cancer therapy prior to BICR-confirmed PD should continue to undergo tumour assessments, and sites must send each scan to the sponsor-designated vendor until PD is confirmed by BICR.

After the primary PFS analysis, central review of scans will no longer be required and investigators will be advised when to stop sending copies of the scans to the CRO conducting the central review.

Further details of the BICR will be documented in the BICR Charter.

3.2 Outcome Variables

3.2.1 Progression free survival (PFS)

PFS is defined as the time from randomisation until the date of objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to progression (i.e. date of PFS event or censoring – date of randomisation + 1). Patients 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 assessment. However, if the patient progresses or dies after two or more missed visits, the patient will be censored at the time of the latest evaluable RECIST 1.1 assessment prior to the two missed visits. The primary PFS endpoint will be derived based on the BICR visits response assessments. PFS will also be derived based on the investigator assessed data for a sensitivity analysis.

Given the scheduled visit assessment scheme (i.e. six-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 288 (i.e. week 41) then two missing visits will equate to 14 weeks since the previous RECIST assessment, allowing for early and late visits (i.e. 2 x 6 weeks + 1 week for an early assessment + 1 week for a late assessment = 14 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from six-weekly to twelve-weekly this will equate to 20 weeks (i.e. take the average of 6 and 12 weeks which gives 9 weeks and then apply same rationale, hence 2 x 9 weeks + 1 week for an early assessment + 1 week for a late assessment = 20 weeks). The time period for the previous RECIST assessment will be from study days 288 to 344 (i.e. week 41 to week 49). From week 49 onwards (when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 26 weeks (i.e. 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks).

If the patient has no evaluable visits or does not have baseline data they will be censored at Day 1 unless they die within two visits of baseline (12 weeks plus 1 week allowing for a late assessment within the visit window).

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

- For BICR assessments, the date of progression will be determined based on the **earliest** of the scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of the reviewer who read baseline first if there is no adjudication for BICR data.
- For investigational assessments, the date of progression will be determined based on the **earliest** of the dates of the component that triggered the progression.
- For both BICR and investigational assessments, when censoring a patient for PFS the patient will be censored at the **latest** of the dates contributing to a particular overall visit assessment.

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

3.2.2 Overall survival (OS)

Overall survival is defined as the time from the date of randomisation until death due to any cause regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy (i.e. date of death or censoring – date of randomisation + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Note: Survival calls will be made within 5 days following the date of data cut-off (DCO) for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the final OS analysis should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly-available death registries. In the event that the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

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

• Admission and discharge dates of hospitalisation

• Study treatment date

• End of treatment date

• Laboratory test dates

• Date of vital signs

• Disease assessment dates on RECIST CRF

• Start and stop dates of alternative anticancer treatment

• Date last known alive on survival status CRF

• End of study date

3.2.3 Objective response rate (ORR)

ORR is defined as the percentage of patients with at least one visit response of CR or PR.

ORR will be calculated using the BICR data to define a visit response of CR or PR, with the denominator defined as subset of all randomised patients with measurable disease at baseline per BICR.

ORR will be calculated using the investigator data as at least one investigator-assessed visit response of CR or PR, with the denominator defined as a subset of all randomised patients with measurable disease at baseline per the site investigator.

Data obtained up until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR. Patients who discontinue randomised treatment without progression, receive a subsequent anti-cancer therapy (note that for this analysis radiotherapy is not considered a subsequent anti-cancer therapy) and then respond will not be included as responders in the ORR.

3.2.4 Duration of response (DoR)

DoR will be defined both based on the BICR data and also based on the site investigator data as the time from the date of first documented response 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 response of PR or CR.

If a patient does not progress following a response, then their DoR will use the PFS censoring time.

3.2.5 Time from randomisation to second progression or death (PFS2)

Time from randomisation to second progression or death (PFS2) is defined as the time from the date of randomisation to the earliest of the progression event (subsequent to that used for the primary variable PFS) or death (i.e. date of PFS2 event or censoring – date of randomisation + 1). The date of the first progression will be programmatically determined from investigator-assessed data. The date of second progression will be recorded by the investigator and defined according to local standard clinical practice. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the electronic case report form (eCRF).

Second progression status should be reviewed at regular assessments following the progression event used for the primary variable PFS (the first progression) and status recorded. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without a second disease progression, i.e. censored at the latest of the PFS or PFS2 assessment date if the patient has not had a second progression or death. However, if the patient experiences a second progression or dies after two or more missed visits, the patient will be censored at the time of the last PFS2 assessment prior to the two missed visits.

3.2.6 Best objective response (BoR)

Best objective response (BoR) is calculated based on the overall visit responses from each RECIST assessment, described in section 3.1.3. It is the best response a patient has had following randomisation, but prior to starting any subsequent cancer therapy and up to and including RECIST progression or the last evaluable assessment in the absence of RECIST progression. Categorisation 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.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 6 weeks minus 1 week, i.e. at least 35 days (to allow for an early assessment within the assessment window), after randomisation. 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 BICR data up until the first progression event. It will also be determined programmatically based on RECIST using all site investigator data up until the first progression event. The denominators for each case 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 ≤7 weeks (i.e. 6 weeks + 1 week to allow for a late assessment within the assessment window) after randomisation, then BoR will be assigned to the progression (PD) category. For patients who die with no evaluable RECIST assessments, if the death occurs >7 weeks after randomisation then BoR will be assigned to the NE category.

A patient will be classified as a responder if the RECIST criteria for a CR or PR are satisfied at any time following randomisation, prior to RECIST progression and prior to starting any subsequent cancer therapy.

3.2.7 Change in TL tumour size

3.2.7.1 General guidance

Tumour shrinkage will be assessed using RECIST TL measurements taken at baseline and at all post baseline assessments prior to progression or start of subsequent anticancer therapy. Tumour size is the sum of the longest diameters of the TLs. TLs are measurable tumour lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to randomisation. The absolute change and percentage change from baseline in TL tumour size at each assessment will be calculated. The percentage change from baseline will be obtained for each patient and post baseline visit by taking the difference between the sum of the TLs at the post baseline visit, week x, and the sum of the target lesions at baseline divided by the sum of the TLs at baseline times 100 (i.e. (week x - baseline) / baseline * 100). The best change in tumour size will then be calculated. More details on TLs and measurements can be found in Appendix F of the protocol.

The primary analysis of tumour shrinkage will be based on BICR; tumour shrinkage per investigator review will be a sensitivity analysis.

3.2.7.2 Missing data imputation methods

For best percentage change in Tumour Size

-TL imputation
The scaling up rule described in section 3.1.1 will be applied to all assessments. All assessments should be included in the derivation of the best percentage change from baseline, including unscheduled assessments.

-Selection of the best percentage change for each patient
This is the biggest decrease or the smallest increase in TL tumour size from baseline. If there are patients with very early (i.e. in week 1 or 2) unscheduled scans, omitting these measurements will be considered.

3.2.8 Disease control rate (DCR)

Disease control rate at 24 weeks (6 months) is defined as the percentage of patients who have a BoR of CR or PR in the first 25 weeks (to allow for a late assessment within the assessment window) or who have SD for at least 23 weeks after start of treatment (to allow for an early assessment within the assessment window). Disease control rate at 6 months and 12 months will be primarily based on BICR visit responses; DCR by investigator assessment of visit responses will be included as a sensitivity analysis.

3.2.9 Patient Reported Outcome (PRO) Endpoints

Patient reported outcomes will be assessed using the FACIT-F, FSKI-19 and CTCAE-PRO questionnaires. Higher scores on the global health status and functioning scales indicate better health status/function. Higher scores on the symptoms scales indicate greater symptom burden.

3.2.9.1 FACIT-F

The validated Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) (now in Version 4) is a 40-item instrument. It comprises the 4 domains from the FACT-G (Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being) as well as a 13-item 'additional concerns' domain that captures information about fatigue.

3.2.9.2 FKSI-19

The Functional Assessment of Cancer Therapy Kidney Symptom Index-19 (FKSI-19) is a validated instrument designed to accurately assess patient self-reported symptom burden to determine treatment impact and evaluate clinical benefit in patients with renal cancer. The instrument includes 19 items covering 4 subscales: Disease-Related Symptoms Subscale – Physical, Disease-Related Symptoms Subscale – Emotional, Treatment Side Effects Subscale, and the Function and Well-Being Subscale. Responses are reported as they apply to the past 7 days. Higher scores indicate better functioning or well-being. The scores for each subscale and the total will be summarised by visit and for change from baseline at each visit.

3.2.9.3 EQ-5D-5L

The EQ-5D is a standardised measure of health status developed by the European Quality of Life (EuroQoL) Group in order to provide a simple, generic measure of health for economic appraisal. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the economic evaluation of health care.

The questionnaire assesses 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 response options (no problems, slight problems, moderate problems, severe problems, and extreme problems) that reflect increasing levels of difficulty. The patient will be asked to indicate his/her current health state by selecting the

most appropriate level in each of the 5 dimensions. The questionnaire also includes a visual analogue scale (VAS), where the patient will be asked to rate their current health status on a scale of 0 to 100, with zero being the worst imaginable health state.

3.2.9.4 PRO CTCAE

The Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) system has been developed by the National Cancer Institute (NCI). PRO-CTCAE is an item-bank of symptoms experienced by patients while undergoing treatment of their cancer. PRO-CTCAE will only be administered in the US, Italy, France, and South Korea where a linguistically-validated version exists. The items investigated are provided in Appendix I of the CSP. PRO CTCAE data will be listed only.

3.2.9.5 Compliance

Summary measures of overall compliance and compliance over time will be derived for each PRO, respectively. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.
- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under PRO follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.
- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing.
- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all time points, divided by total number of questionnaires expected to be received across all time points multiplied by 100.
- Overall patient compliance rate is defined for each randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the time point (as defined above), divided by number of patients still expected to complete questionnaires. Similarly, the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires.

3.3 Safety

3.3.1 General considerations for safety and PRO assessments

Time windows will need defining for any presentations that summarise 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 summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit.
- All unscheduled visit data should have the potential to be included in the summaries.

- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day. For example, the visit windows for vital signs data (with 2 weeks between scheduled assessments for the first two cycles) are:
  - Day 15, visit window 2 – 21
  - Day 29, visit window 22 – 35
  - Day 43, visit window 36 – 49
  - Day 57, visit window 50 – 63

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

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

- For visit based summaries

  If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be summarised, or the earlier, in the event the values are equidistant from the nominal visit date. When there are multiple values on the same day that are eligible to be summarised based on the above criteria then a two-step approach will be taken: 1) If the value is strictly numeric then the average will be taken 2) If the value is strictly character (e.g. NEGATIVE, +++) then for post-baseline visits the worst record will be selected and for baseline the best record as this is the most conservative. The listings should highlight the value for the patient that contributed to the summary table, wherever feasible. Note: in summaries of extreme values all post baseline values collected are used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date.

  To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group, visit data should only be

summarised if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.

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

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

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

For laboratory data the following applies:
Numerical summaries will provide the mean, standard deviation, median, minimum, maximum, and lower and upper quartile for visit based tabular summaries.

3.3.2 Handling of partial dates

For missing start dates for AEs and concomitant medications/procedures, the following will be applied:

- Missing day - Impute the 1st of the month unless month is same as month of first dose of study drug then impute first dose date
- Missing day and month – impute 1st January unless year is the same as first dose date then impute last dose date.
- Completely missing date – impute first dose date unless the end date is less than the first dose date, in which case impute the 1st January of the same year as the end date.

When imputing a start date ensure that the new imputed date is sensible e.g. is prior to the end date of the AE.

For missing stop dates of AEs or concomitant medications/procedures, the following will be applied:

- Missing day - Impute the last day of the month unless month is the same as month of the last dose of study drug then impute last dose date.
- Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date.
- Completely missing date – do not impute.

The imputation of dates will be used to decide if an observation is treatment emergent for adverse events or concomitant medications. The imputed dates are not used to calculate durations.

3.3.3 Exposure and dose interruptions

Exposure will be defined as follows:

Total (or intended) exposure of study treatment

- Total (or intended) exposure = min(last dose date where dose > 0 [units], date of death, date of DCO) – first dose date +1

Actual exposure of study treatment

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

The actual exposure calculation makes no adjustment for any dose reductions that may have occurred. Sunitinib will be given as two 25 mg capsules PO daily, with or without food for 4 weeks on/ 2 weeks off. The two week break in dosing for sunitinib will not be considered a dose interruption for the calculation of actual exposure.

Exposure will also be measured by the number of cycles received. A cycle corresponds to a period of 42 days. If a cycle is prolonged due to toxicity, this should still be counted as one cycle. A cycle will be counted if treatment is started even if the full dose is not delivered.

Estimated treatment compliance will be derived from the actual administration days (accounting for planned dose interruptions), divided by the total planned administration days (last dose date - first dose date + 1). Patients who received no study treatment will have zero compliance.

Missed or forgotten doses

Missed and forgotten doses should be recorded on the EX and EX1 modules for sunitinib and savolitinib respectively as a dose interruption with the reason recorded as “Subject forgot to take dose”. These missed or forgotten doses will not be included as dose interruptions in the

summary tables but the information will appear in the listing for dosing. However, these missed and forgotten doses will be considered in the derivation of actual exposure.

Patients who permanently discontinue during a dose interruption

If a patient permanently discontinues study treatment during a dose interruption, then the date of last administration of study medication recorded on DOSDISC will be used in the programming.

Safety Follow-up

Total Safety Follow-up = min((last dose date + 30 days), date of withdrawal of consent, date of death, date of DCO, date of first dose of subsequent anti-cancer therapy) – first dose date +1.

3.3.4 Dose intensity

Relative dose intensity (RDI) is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. Relative dose intensity (RDI) will be defined as follows:

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

3.3.5 Adverse Events

AEs and SAEs will be collected throughout the study, from date of informed consent until 30 days after the last dose of study treatment. Events will be defined as treatment emergent if they onset, or worsen (by investigator report of a change in intensity), during the treatment period or 30 day safety follow up period as defined in the protocol. The Medical Dictionary for Regulatory Activities (MedDRA) (using the latest or current MedDRA version) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 4.03).

AEs of special interest

Some clinical concepts (including some selected individual preferred terms and higher-level terms) have been considered “AEs of special interest” (AESI) to the savolitinib program. AESIs represent pre-specified risks that are considered to be of importance to a clinical development program.

These AESIs have been identified as a list of categories provided by the patient safety team. Other categories may be added as necessary or existing terms may be merged. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AEs of interest and identified which higher-level terms and which preferred terms contribute to each AESI. Further reviews may take place prior to database lock (DBL) to ensure any further terms not already included are captured within the categories.

3.4 Pharmacokinetic Variables

Pharmacokinetic concentration data will be collected as per the protocol.

PK parameters for savolitinib will be estimated from the plasma concentration data from the sparse PK sampling regimen employed by combining with data from other studies and use of population PK modelling techniques. These data, dependent on the timing of the work, may be reported outside the CSR.

4 ANALYSIS METHODS

4.1 General Principles

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

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

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

- For continuous data, the mean, median, upper and lower quartiles 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 statistical analysis, confidence intervals should be rounded to 2 decimal places and p-values to 3 decimal places.

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

- SAS® version 9.4 will be used for all analyses.

In general, for efficacy and PRO endpoints the last observed measurement prior to randomisation will be considered the baseline measurement. However, for PRO endpoints if an evaluable assessment is only available after randomisation but before the first dose of randomised treatment then this assessment will be used as baseline. For safety endpoints the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The percentage change from baseline will be calculated as (post-baseline value - baseline value) / baseline value x 100.

PROs, haematology, clinical chemistry, vital signs and ECGs will be summarised at end of treatment. If a scheduled end of treatment visit occurs then this visit will be summarised, otherwise the first visit post discontinuation of IP will be used.

Efficacy and PRO data will be summarised based upon the FAS. Safety and treatment exposure data (apart from treatment compliance) will be summarised based upon the safety analysis set. Study population, demography and treatment compliance data will be summarised based upon the FAS.

4.2 Analysis Methods

4.2.1 Time-to-event endpoint considerations

PFS and OS will be analysed using a log-rank test stratified by IMDC risk group (poor, intermediate, favourable), and line of therapy (1st line; previously treated with VEGF TKI; previously treated without VEGF TKI) providing there are at least 10 PFS events within each stratum, using the Breslow approach for handling ties (Breslow, 1974). If there are not at least 10 PFS events within each stratum, then the log-rank test will be stratified by IMDC risk group only, providing there are at least 10 PFS events within each stratum. If not, then the log-rank test will be unstratified.
The covariates in the statistical modelling will be based on the values entered into IVRS at randomisation, even if it is subsequently discovered that these values were incorrect.

Proportionality assumption

The assumption of proportionality may be assessed in suspicion of non-proportionality is raised by the PFS KM curves. Proportional hazards would then be tested firstly by examining plots of complementary log-log (event times) versus log (time) and, if these raise concerns, by fitting a time dependent covariate to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect can be described by presenting piecewise HR calculated over distinct time-periods. In such circumstances, the HR from the primary analysis can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found this may be a result of a treatment-by-covariate interaction, which will be investigated.

4.2.2 Progression free survival (PFS)

PFS will be analysed using a stratified log-rank test as described in section 4.2.1. The effect of treatment will be estimated by the HR together with its corresponding 95% CI from a sensitivity analysis using a stratified Cox proportional hazards model with treatment as a covariate (ties=Efron), where the stratification is as specified for the log-rank test. Kaplan-

Meier plots will be presented by treatment arm. Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent will be provided along with the median and 95% CI for the median (using Kaplan-Meier technique) for each treatment. The progression-free survival rate will also be presented along with the 95% CI at 3, 6, 9 and 12 months.

The treatment status at progression of patients at the time of analysis will be summarised. This will include the number (%) of patients who were on treatment at the time of progression, the number (%) of patients who discontinued study treatment prior to progression, the number (%) of patients who have not progressed and were on treatment or discontinued treatment. This will also provide distribution of number of days prior to progression for the patients who have discontinued treatment.

The primary analysis of PFS will be based on BICR.

4.2.2.1 PFS sensitivity analyses

A sensitivity analysis for PFS based on investigator review will be performed.
Additionally, a sensitivity analysis for the subset of subjects with centrally confirmed MET by SAVOIR Clinical Trial Assay (CTA) will also be conducted.
Deviation bias may be assessed by repeating the PFS analysis excluding patients with deviations that may affect the efficacy of trial therapy, as described in section 2.2.

Additional supportive summaries
The duration of follow-up will be summarised using median time from randomisation to date of censoring (date last known to have not progressed) in censored (not progressed) patients only, presented by treatment group.
All of the collected RECIST 1.1 data will be listed for all randomised patients. In addition, a summary of new lesions (i.e. sites of new lesions) may be produced.

4.2.3 Overall survival (OS)

OS will be analysed using a stratified log-rank test as described in section 4.2.1.

The effect of treatment will be estimated by the HR together with its corresponding 95% CI from a sensitivity analysis using a stratified Cox proportional hazards model with treatment as a covariate (ties=Efron), where the stratification is as specified for the log-rank test. Kaplan-Meier plots will be presented by treatment arm. Summaries of the number and percentage of patients who have died, those still in survival follow-up, those alive at year 1, those lost to follow-up and those who have withdrawn consent will be provided along with the median and 95% CI for the median (using Kaplan-Meier technique) for each treatment.

The number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if their survival status was not defined at the DCO.

4.2.4 Objective response rate (ORR)

The ORR will be based on the BICR RECIST data, and using all scans regardless of whether they were scheduled or not.
Summaries will be produced that present the number and percentage of patients with a tumour response of CR or PR. Summaries of ORR will be based on BICR; ORR per investigator review will be a sensitivity analysis.

For each treatment arm, best objective response (BoR) will be summarised by n (%) for each category (CR, PR, SD, PD and NE). This will be presented based on BICR; BoR per investigator review will be presented separately. No formal statistical analyses are planned for BoR.

4.2.4.1 Confidence intervals for response rates

The confidence intervals for the response rate within a treatment will be calculated using the Clopper-Pearson method.

The Clopper-Pearson interval comprises all θ for which precisely computed, 'exact' aggregate tail areas are not less than $\frac{\alpha}{2}$ if we wish to obtain the $(1-\alpha)\%$ confidence interval for p. Numerical values are obtained iteratively, or by use of published tables that rely on equivalent F distributions, or, according to Brown et al (2001), the lower endpoint is the α/2 quantile of a beta distribution Beta(r, n − r + 1), and the upper endpoint is the 1 − α/2 quantile of a beta distribution Beta(r + 1,n – r). For the iterative approach, the interval is$[L,U]$, with $L \le p \le U$, such that for all θ in the interval:

(i) if $L \le \theta \le p$: $$p_r + \sum_{j:r<j\le n} p_j \ge \frac{\alpha}{2}$$

(ii) if $p \le \theta \le U$: $$\sum_{j:0\le j<r} p_j + p_r \ge \frac{\alpha}{2}$$

respectively, where:

$$p_j = P[R=j] = \binom{n}{j}\theta^j(1-\theta)^{n-j}$$

$$j = 0,1,...n$$

Where n is the number of observations, r is the observed number of successes and p is the proportion of successes.

4.2.5 Duration of response (DoR)

Descriptive data will be provided for the duration of response in responding patients, including the associated Kaplan-Meier curves (without any formal comparison or p-value attached), medians and 95% confidence interval.

Summaries of DoR will be based on BICR; DoR per investigator review will be a sensitivity analysis.

4.2.6 Time from randomisation to second progression or death (PFS2)

Analyses on time from randomisation to second progression or death (PFS2) will not be conducted.

4.2.7 Change in TL tumour size

The best change in TL tumour size from baseline, (where best change in TL size is the maximum reduction from baseline or the minimum increase from baseline in the absence of a reduction) will be summarised by randomised treatment group. This will be presented in a waterfall plot (bar chart) indicating the best percentage change from baseline in the sum of the diameters of TLs based on BICR. Reference lines at the +20% and 30% change in tumour size levels will be added to the plots, which correspond with the definitions of progression and 'complete or partial' response respectively.

4.2.8 Disease control rate (DCR)

DCR at 6 and 12 months will be presented for each treatment group together with 95% exact (Clopper Pearson) confidence intervals.

Summaries of DCR will be based on BICR; DCR per investigator review will be a sensitivity analysis.

4.2.9 Subgroup analyses

No subgroup analysis will be performed.

4.2.10 Patient reported outcomes (PROs)

Summaries of the FACIT-F, FKSI-19 and EQ-5D-5L will be based on the instruments' respective scoring manuals.

The EQ-5D profile will be converted into the EQ-5D index by applying the UK EQ-5D-5L value set (Devlin et al. 2017).

Descriptive statistics will be calculated for each scheduled visit/time point in the study for each trial arm. These will report the number of patients, the number of questionnaires

completed at each visit, the number and the proportion responding to each dimension/domain/subscale. Summary statistics will be reported for the total FACIT-F score, total FKSI-19 score, EQ-5D index score and the EQ-5D VAS score, and the change from baseline for the total FACIT-F score, total FKSI-19 score, EQ-5D index score and the EQ-5D VAS score. For EQ-5D-5L only, summary statistics will be presented by dimension and visit.

Summary measures of overall compliance and compliance over time will be derived for the FACIT-F, FKSI-19 and EQ-5D-5L respectively. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.
- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under HRQoL follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.
- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing.
- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all time points, divided by total number of questionnaires expected to be received across all time points multiplied by 100.
- Overall patient compliance rate is defined for each randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by 100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the time point (as defined above), divided by number of patients still expected to complete questionnaires. Similarly the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires.
PRO-CTCAE will be listed over time.

4.2.11 Patient disposition

Patient disposition will be summarised using the all patients analysis set. The number of patients who were enrolled will be summarised. The number and percentage of patients within each treatment group will be presented by the following categories; randomised, not randomised (and reasons), full analysis set, randomised who received study treatment, randomised who did not receive study treatment (and reasons), patients ongoing study treatment at DCO, patients who were not ongoing treatment at DCO (and reasons), patients ongoing in the study and patients who terminated the study (and reasons).

A separate table will present the number and percentage of patients randomised to each treatment group by country and centre. This table will be based on the full analysis set.

The number of patients assigned to each analysis set and excluded from each analysis set (and reasons) will also be summarised for each treatment group.

4.2.12 Demography data, patient and disease characteristics

Demography data for age, age group (<65; >=65 years), sex, race, ethnic group and country will be summarised by treatment group for the full analysis set. Patient characteristics at baseline will be summarised by treatment for the full analysis set. These include height, weight, weight group (<50kg; >=50kg), BMI and BMI group (<25; 25-30; >30). Subgroup characteristics at baseline will also be summarised by treatment for the full analysis set. These include IDMC risk group from IVRS (poor; intermediate; favourable), line of therapy (1st line; >= 2nd line with prior VEGF TKI; >= 2nd line without prior VEGF TKI), subtype of MET-driven criteria (MET amp; HGF amp; MET mutations; Chromosome 7 gain), histology type (Papillary carcinoma; Papillary carcinoma, other variant; Other), Histology subtype (Type 1; Type 2; Unspecified) and Karnofsky Performance Status (100%; 90%; 80%).

Disease characteristics at baseline will be summarised by treatment group for the full analysis set. These include primary tumour location, primary tumour TNM classification, regional lymph nodes TNM classification, distant metastases TNM classification, tumour grade, AJCC Staging, time from diagnosis to randomisation (<=6 months; >6 months) and overall disease classification. TNM characteristics will be taken from the time of diagnosis, with all other disease characteristics taken from the most recent tumour sampling.

The extent of disease at baseline will also be summarised by treatment group for the full analysis set. This will summarise the extent of disease (locally advanced; metastatic) by the site of disease (including total).

Medical and surgical histories will be summarised by treatment group. MedDRA preferred term (PT) will be summarised within MedDRA system organ class (SOC).

Nicotine use and consumption at baseline will be summarised by treatment group for the full analysis set. This will summarise the number and percent of patients who were never, former or current smokers as well as summary statistics for the number of pack years. This information will be summarised from the data collected at Part 1: Screening as both pack years and smoking history are collected at this timepoint.

All summaries will also include rows for missing data where appropriate.

4.2.13 Concomitant medication

The number and percentage of patients who take allowed concomitant medications, and those who take disallowed concomitant medications during the study, will be presented by treatment group. Generic term will be summarised within ATC classification, classified according to the WHODrug dictionary. The number of regimens of previous anti-cancer therapy at baseline will be summarised categorically and by summary statistics by treatment group. Previous disease-related treatment modalities and post-discontinuation disease related anti-cancer therapy will also be summarised by treatment group.

4.2.14 Safety

Safety and tolerability data will be presented using summaries and descriptive statistics. The safety analysis set will be used for all safety and tolerability tables, figures and listings except where expressly noted.

For change from baseline summaries for vital signs, laboratory data, ECG, ECHO/MUGA, the baseline value will be the latest result obtained prior to the start of investigational product.

The denominator used in laboratory summaries will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality. For example:

- If a CTCAE criterion involves a change from baseline, evaluable patients would have both a pre-dose and at least 1 post-dose value recorded.
- If a CTCAE criterion does not consider changes from baseline, to be evaluable the patient need only have 1 post dose-value recorded.

The denominator in vital signs data should include only those patients with recorded data.

The number of hospitalisations, type of attendance, reason, and time spent in hospital (summed across all patients) will be summarised by arm. Additionally the number of ICU/HDU stays, type of attendance, reason, and time spent in ICU/HUD (summed across all patients) will also be summarised by arm.

4.2.14.1 Exposure

The following summaries will be produced based on the safety analysis set:

- Summary of duration of exposure of savolitinib and sunitinib. This will include the total treatment duration, actual treatment duration and duration of therapy at the starting dose.
- RDI of savolitinib and sunitinib (see section 3.3.4)
- Summary of interruptions and reductions of savolitinib and sunitinib

Study treatment compliance will be presented for the full analysis set. Patients who receive no study treatment will have zero compliance.

4.2.14.2 Adverse events (AEs)

All AE data will be listed individually by patient. Any AE occurring before treatment with either IPs will be included in the data listings and flagged but will not be included in the summary tables of AEs.

A separate summary of AEs (where reported) occurring more than 30 days after discontinuation of both IPs, as well as those occurring prior to treatment will be produced. These events will not be included in any other AE summaries.

An overview table by study treatment will summarise the number and percentage of patients with at least one of the following AEs where patients with more than one AE in a particular category are counted only once in that category:

- Any AE
- Any AE causally related to treatment
- Any AE of CTCAE grade 3 or higher
- Any AE of CTCAE grade 3 or higher, causally related to treatment
- Any AE with outcome = death
- Any AE with outcome = death, causally related to treatment
- Any SAE (including events with outcome=death)
- Any SAE (including events with outcome=death), causally related to treatment
- Any SAE leading to discontinuation of treatment
- Any AE leading to discontinuation of treatment
- Any AE leading to dose modification
- Any AE leading to dose reduction
- Any AE leading to dose interruption
- Any other significant AE (if applicable)

The number and percentage patients reporting each AE will be summarised, by SOC (sorted alphabetically) and PT (sorted by descending overall total), with grouped medical topics underneath. The following summaries will be produced:

- AEs by SOC and PT.
- AEs by maximum reported CTCAE grade, by SOC and PT.
- AEs of CTCAE grade 3 or higher, by SOC and PT.
- AEs causally related to treatment, by SOC and PT.
- AEs with outcome of death, by SOC and PT.
- AEs with outcome of death, causally related to treatment, by SOC and PT.
- SAEs by SOC and PT.
- SAEs causally related to treatment, by SOC and PT.

- AEs leading to discontinuation of treatment, by SOC and PT.

In the above summaries, patients with more than one AE within a particular SOC are counted only once for that SOC. Similarly, patients with more than one AE within a particular PT are counted only once for that PT. For summaries by maximum reported CTCAE grade with multiple AEs within a particular SOC or PT will be counted under the category of their most severe AE within that SOC or PT.

Every CTC grade change within an adverse event episode will be captured during the study. Changes in CTC grades in AEs will not be reported in the CSR, but may be reported separately.

A summary of deaths will be provided, detailing deaths due to disease under investigation only, deaths where AE outcome is death only, deaths where AE outcome is death only (AE start date falling after 30 day follow up period), AE where death is related to disease and AE with outcome of death and other deaths (deaths that are not captured in the previous categories).

4.2.14.3 Adverse events of special interest (AESI)

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

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

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

4.2.14.4 Laboratory evaluations

For clinical chemistry and haematology, numerical summaries of absolute values and change from baseline will be generated which include the mean, standard deviation, median, minimum value, maximum value, and lower and upper quartiles at each visit in which laboratory values are taken. Shift tables will also be produced. Plots of both the maximum post baseline ALT and AST versus the maximum post-baseline total bilirubin, expressed as multiples of their upper limit of reference range will be produced.

As a primary interpretation of potential drug induced liver injury, outputs related to patients with liver function tests (LFTs) that meet biochemical criteria for potential Hy's Law and with elevations of LFTs will be produced. This will include all patients who have ALT or AST ≥ 3x ULN and total bilirubin ≥ 1.5x ULN, at any time during the study. Biochemical criteria for potential Hy's Law are defined as any situation where a study patient has an increase in both AST or ALT ≥ 3×ULN and TBL ≥ 2×ULN, irrespective of ALP, at any point during the study. The elevations do not have to occur at the same time or within a specified time frame.

For all patients who meet the criteria for potential Hy's Law, a plot of AST, ALT, ALP and total bilirubin over time will be produced. Additionally the liver diagnostic, liver risk factor, and sign or symptom will be summarised for each patient (with potential Hy's Law) by time point.

4.2.14.5 Vital signs

Summaries of vital signs data will include all data obtained up until 30 days after the last dose of study treatment. Absolute values and change from baseline for diastolic and systolic BP, heart rate, weight and body temperature will be summarized at each visit by treatment.

4.2.14.6 Electrocardiograms, ECHO/MUGA

Summaries of ECG and ECHO/MUGA data will include all data obtained up until end of treatment. Absolute values and change from baseline for corrected QC interval (QTc) and left ventricular ejection fraction (LVEF) will be summarized at each visit by treatment. For QTc, there are 3 recordings at each visit and therefore for the analysis the average of the non-missing values will be taken as the result at that visit.

4.2.14.7 Urinalysis

Any qualitative safety assessments for urinalysis will be presented as is deemed appropriate. This may include, but is not restricted to, presentation of parameters against time, concentration, or shift plots. Appropriate scatter plots will also be considered to investigate trends in parameters compared with baseline.

4.2.15 Pharmacokinetic data

The plasma concentration-time data will be analysed by non-linear mixed effects modelling in order to evaluate the pharmacokinetic characteristics of savolitinib, quantify variability in the pharmacokinetics, identify demographic or pathophysiological covariates, which may explain the observed variability and explore exposure-response relationships. The savolitinib plasma concentration data obtained from the samples collected in this study will be pooled with data from other studies in order to perform a population PK analysis. The results of this analysis will be reported in a separate population PK report.

For the CSR, listings of plasma concentration of savolitinib and metabolites, M2 and M3 will be presented.

4.2.16 Exploratory Endpoints

As explained in section 1.1.4 the following exploratory outcome measures: [redacted] will be reported in another CSR so no detailed analysis methods are given in this section.

5 INTERIM ANALYSES

An interim futility analysis of PFS was planned when a total of approximately 36 PFS events confirmed by BICR hadobserved (30% of 121 PFS events, which was estimated to occur 17 months after the first patient had been randomised).

Following IDMC endorsement to terminate recruitment into SAVOIR, the interim analysis will not occur.

However interim reviews of safety data by IDMC will continue until the end of the study.

IDMC

This study will use an external Independent Data Monitoring Committee (IDMC) to perform interim reviews of accumulating study safety data. This committee will be composed of therapeutic area experts, including an independent hepatic expert to review hepatic – specific outputs and liver laboratory data, and a statistician, who are not employed by AZ, and do not have any major conflict of interest. Following the interim analysis the IDMC will recommend whether the study should continue unchanged, be terminated, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca. The report will only include the recommendation and any potential protocol amendments. A separate IDMC charter will be developed which will contain details of the IDMC members and clearly define the responsibilities of the IDMC. The charter will describe the plan to provide safety reports (individual case safety report or serious adverse event report) and follow up safety reports concerning liver dysfunction to the IDMC in real time.

In addition to the periodic review of safety data by an IDMC, the safety of all AstraZeneca clinical studies is closely monitored on an on-going basis by AstraZeneca representatives in consultation with the Patient Safety Department. Issues identified will be addressed; this could involve, for instance, amendments to the study protocol and letters to investigators. Abnormalities in liver function tests as well as any clinical adverse events associated with liver dysfunction will be reported quarterly to the FDA.

Full details of the IDMC procedures and processes can be found in the IDMC Charter.

6 CHANGES OF ANALYSIS FROM PROTOCOL

In a change from the protocol (v6) OS will be analysed using a stratified log-rank test as described in section 4.2.1 and the effect of treatment will be estimated by the HR together with its corresponding 95% CI from a sensitivity analysis using a stratified Cox proportional hazards model as described for PFS in section 4.2.2.
The summaries and Kaplan Meier plots for PFS2 referenced in the protocol (v6) will be not produced due to the reduced sample size and low anticipated maturity of the data.

A summary of the subtypes of MET-driven by SAVOIR CTA will be presented with the patient characteristics and the PFS analyses will be repeated in patients with central SAVOIR CTA results. The rational for the inclusion of this analyses is that patient inclusion criteria to study was “MET-driven” status (as per the primary objective) and the only “true” result for this is only obtained by the SAVOIR Clinical Trial Assay (CTA). Therefore, it is of importance to produce PFS for these confirmed “MET-driven” patients.

7 REFERENCES

Armstrong et al 2016
Armstrong AJ, Halabi S, Eisen T, Broderick S, Stadler WM, Jones RJ, et al. Everolimus versus sunitinib for patients with metastatic non-clear cell renal cell carcinoma (ASPEN): a multi-centre, openlabel, randomised phase 2 trial. Lancet Oncol. 2016 Mar;17(3):378-88.

Basch et al 2009
Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009 Dec 2;101(23):1624-32

Bellmunt and Dutcher 2013
Bellmunt J, Dutcher J. Targeted therapies and the treatment of non-clear cell renal cell carcinoma. Ann Oncol. 2013 Jul;24(7):1730-40.

Breslow, NE. 1974
Covariance Analysis of Censored Survival Data. Biometrics. 1974; 30:89–99

Brown LD, Cai TT, DasGupta A.
Interval estimation for a binomial proportion. Statist Sci. 2001; 16:101–133

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med. 2009 Feb 28;28(5):739-61.

Choueiri et al 2013
Choueiri TK, Vaishampayan U, Rosenberg JE, Logan TF, Harzstark AL, Bukowski RM, et al. Phase II and biomarker study of the dual MET/VEGFR2 inhibitor foretinib in patients with papillary renal cell carcinoma. J Clin Oncol. 2013 Jan 10;31(2):181-6.

Clopper and Pearson, 1934
Clopper C and Pearson E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934; 26: 404-413.

Devlin et al. 2017
Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of Life: An EQ-5D-5L Value Set for England. Health Economics. 2017;1-16

Escudier et al 2014
Escudier B, Porta C, Schmidinger M, Algaba F, Patard JJ, Khoo V, et al. ESMO Guidelines Working Group. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014 Sep;25 Suppl 3:iii49-56.

Fischer et al 1998
Fischer J, Palmedo G, von Knobloch R, Bugert P, Prayer-Galetti T, Pagaono F, et al. Duplication and overexpression of the mutant allele of the MET proto-oncogene in multiple hereditary papillary renal cell tumours. Oncogene. 1998 Aug 13;17(6):733-9.

Gail M, Simon R.
Testing for qualitative interactions between treatment effects and subject subsets. Biometrics. 1985; 41:361-72

Gan et al 2015
Gan HK, Lickliter J, Millward M, Gu Y, Weiguo S, Qi C, et al. cMet: Results in papillary renal cell carcinoma of a phase I study of AZD6094/volitinib leading to a phase 2 clinical trial with AZD6094/volitinib in patients with advanced papillary renal cell cancer (PRCC). J Clin Oncol 33, 2015 (suppl 7; abstr 487).

Grubb et al 2007
Grubb RL 3rd, Franks ME, Toro J, Middelton L, Choyke L, Fowler S, et al. Hereditary leiomyomatosis and renal cell cancer: a syndrome associated with an aggressive form of inherited renal cancer. J Urol. 2007 Jun;177 (6):2074-9; discussion 2079-80.

Jeffers et al 1997
Jeffers M, Schmidt L, Nakaigawa N, Webb CP, Weirich G, Kishida T et al. Activating mutations for the met tyrosine kinase receptor in human cancer. Proc Natl Acad Sci USA. 1997 Oct 14;94(21):11445-50

Klatte et al 2009
Klatte T, Pantuck AJ, Said, JW, Seligson DB, Rao NP, La Rochelle JC et al. Cytogenetic and molecular tumour profiling for type 1 and type 2 papillary renal cell carcinoma. Clin Canc Res 2009; Feb 15;15(4):1162-9.

Kovacs 1993
Kovacs G. Molecular cytogenetics of renal cell tumours. Advanced Can Res 1993; 62: 89-124.

Kroeger et al 2013
Kroeger N, Xie W, Lee JL, Bjarnason GA, Knox JJ, Mackenzie MJ, et al. Metastatic nonclear cell renal cell carcinoma treated with targeted therapy agents: characterization of survival outcome and application of the International mRCC Database Consortium criteria. Cancer. 2013 Aug 15;119(16):2999-3006

Linehan et al 2012
Linehan WM. Genetic basis of kidney cancer: role of genomics for the development of disease-based therapeutics. Genome Res. 2012 Nov;22(11):2089-100.

Litwin et al 1999
Litwin MS, Lubeck DP, Henning JM, et al. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: Results of the CaPSURE database. J Urol. 1998 Jun;159(6):1988-92.

Ljungberg et al 2011
Ljungberg B, Campbel SC, Cho HY, Jacqmin D, Lee JE, Weikert S, et al. The epidemiology of renal cell carcinoma. Euro Urol 2011; 60(4):615-621

Ljungberg et al 2015
Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, et al. EAU guidelines on renal cell carcinoma: 2014 update. Eur Urol. 2015 May;67(5):913-24

Motzer et al 2002
Motzer RJ, Bacik J, Mariani T, Russo P, Mazumdar M, Reuter V. Treatment outcome and survival associated with metastatic renal cell carcinoma of non-clear-cell histology. J Clin Oncol. 2002 May 1;20(9):2376-81.

Motzer et al 2015
Motzer RJ, Hutson TE, Glen H, Michaelson MD, Molina A, Eisen T, et al. Lenvatinib, everolimus, and the combination in patients with metastatic renal cell carcinoma: a randomised, phase 2, open-label, multicentre trial. Lancet Oncol. 2015 Nov;16(15):1473-82

NCCN; v. 2.2017, dated October 31, 2016
National Comprehensive Cancer Network. Kidney Cancer (Version 2.2017; dated October 31, 2016). URL: https://www.nccn.org/professionals/physician_gls/pdf/kidney.pdf

New York Heart Association 1994
The Criteria Committee of the New York Heart Association. Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. 9th ed. Boston, Mass: Little, Brown & Co; 1994:253-256.

Rao et al 2009
Rao D, Butt Z, Rosenbloom S, Robinson D, Von Roenn J, Kuzel TM, et al. A comparison of renal cell carcinoma symptom index (RCC-SI) and the Functional Assessment of Cancer Therapy Kidney Symptom Index (FKSI). J Pain Symptom Manage. 2009 Aug;38(2):291-8.

Schmidt et al 1997
Schmidt L1, Duh FM, Chen F, Kishida T, Glenn G, Choyke P. Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. Nat Genet. 1997 May;16(1):68-73.

Siegel RL, Miller KD, Jemal A. 2016
Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2016. CA Cancer J Clin 2016; 66:7-30

Sprangers et al 1992
Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. J Clin Epidemiol. 2002 Nov;55(11):1130-43

Sun X, Chen C.
Comparison of Finkelstein's Method With the Conventional Approach for Interval-Censored Data Analysis. Stat Biopharm Res. 2010; 2:97-108

Sweeney et al 2002
Sweeney P, El-Naggar AK, Lin SH, Pisters LL. Biological significance of c-met over expression in papillary renal cell carcinoma. J Urol. 2002 Jul;168(1):51-5.

Tannir et al 2015
Tannir NM, Jonasch E, Albiges L, Altinmakas E, Ng CS, Matin SF, et al. Systemic Therapy for Non–clear Cell Renal Cell Carcinomas: Everolimus Versus Sunitinib Prospective Evaluation in Metastatic Non–Clear Cell Renal Cell Carcinoma (ESPN): A Randomized Multicenter Phase 2 Trial. Eur Urol. 2016 May;69(5):866-74.

The Cancer Genome Atlas Research Network 2016
Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, Davis C, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med 2016; 374:135-145.

Van Hout et al, 2012
Van Hout B, Janssen M.F, Feng Y.S, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard A.S. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets: VALUE IN HEALTH 15. 2012; 708 –715

Vera-Badillo et al 2015
Vera-Badillo FE, Templeton AJ, Duran I, Ocana A, de Gouveia P1, Aneja P, et al. Systemic Therapy for Non–clear Cell Renal Cell Carcinomas: A Systematic Review and Meta-analysis. Eur Urol. 2015 Apr;67(4):740-9.