CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR HTA Programme on behalf of NICE

Cabozantinib for untreated locally advanced or metastatic renal cell carcinoma

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors Dr Jonathan Shepherd, Principal Research Fellow
Professor Joanne Lord, Professorial Fellow in Health Economics.
Ms Petra Harris, Research Fellow
Mr Olu Onyimadu, Research Fellow (Health Economics)
Dr Geoff Frampton, Senior Research Fellow
Maria Chorozoglou, Senior Research Fellow (Health Economics)

Correspondence to Dr Jonathan Shepherd
Southampton Health Technology Assessments Centre (SHTAC)
Wessex Institute
Alpha House
Enterprise Road, University of Southampton Science Park
Southampton SO16 7NS
www.southampton.ac.uk/shtac

Date completed: 10th April 2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 17/109/10.

Acknowledgements

We are grateful to the clinical advisors to the ERG:
Professor Robert Hawkins, Cancer Research UK Professor/Honorary Consultant in Medical Oncology, Christie Cancer Hospital Manchester and University of Manchester.
Dr Matthew Wheater, Consultant Medical Oncologist, University Hospital Southampton NHS Foundation Trust.

We would also like to thank: Karen Welch, Information Specialist at SHTAC, for appraising the literature search strategies in the company's submission, running updates of the company's clinical effectiveness searches and searching for ongoing studies; and Dr Emma Loveman Senior Reviewer / Partner, Effective Evidence LLP, for providing a quality assurance review of the draft ERG report.

Declared competing interests

None from the authors. Professor Hawkins declares receipt of speaker fees/honoraria from Ipsen, Novartis, Pfizer and Aveo pharmaceuticals within the previous 12 months. Dr Jenner declares being paid as a speaker for a Pfizer sponsored evening meeting on the topic of managing co-morbidities in metastatic RCC patients (July 2017).

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Shepherd, J., Lord, J., Onyimadu, O., Harris, P., Frampton, G., Chorozoglou, M. Cabozantinib for untreated locally advanced or metastatic renal cell carcinoma: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2018.

Contributions of authors

Jonathan Shepherd critically appraised the clinical effectiveness review, drafted the report, project managed the assessment and is the project guarantor. Joanne Lord critically appraised the economic evaluation, and drafted the report. Petra Harris critically appraised the clinical effectiveness review and drafted the report. Olu Onyimadu critically appraised the health economic review, critically appraised the economic evaluation, and drafted the report. Geoff Frampton critically appraised the clinical effectiveness review and drafted the report. Maria Chorozoglou critically appraised the economic evaluation, and drafted the report.

Word count: 45,992

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Adverse event
AJCC	American Joint Cancer Committee
CABOSUN	Cabozantinib-sunitinib
CHMP	Committee for Medicinal Products for Human Use
CS	Company submission
CSR	Clinical study report
CT	Computerised tomography
DA	Dynamic allocation
DIC	Deviance information criteria
EAU	Updated European Association of Urology
ECOG	Eastern Cooperative Oncology Group
EPAR	European Public Assessment Report
ERG	Evidence Review Group
ESMO	European Society of Medical Oncology
FDA	Food and Drug Administration
FP	Fractional polynomial
HR	Hazard ratio
HRQoL	Health related quality of life
ICER	Incremental cost effectiveness ratio
IMDC	International Metastatic RCC Database Consortium
IRC	Independent radiology committee
ITC	Indirect treatment comparison
ITT	Intention-to-treat
IV	Intravenous
KM	Kaplan-Meier
MET	Hepatocyte growth factor receptor protein
MCMC	Memorial Sloan-Kettering Cancer Center
NCCN	National Comprehensive Cancer Network
NE	Not estimable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
OS	Overall survival
ORR	Objective response rate
PAS	Patient access scheme
PD	Progressed disease
PF	Progression free
PFS	Progression free survival
PH	Proportional hazards
PPES	Palmar-plantar erythrodysaesthesia syndrome
QALY	Quality adjusted life year
QoL	Quality of life
RCC	Renal cell carcinoma
RTKs	Receptor tyrosine kinases
SLR	Systematic literature review
SmPC	Summary of Product Characteristics
TEAE	Treatment emergent adverse events
TNM	Tumour Node Metastasis
TTD	Time to treatment discontinuation
VEGF	Vascular endothelial growth factor (VEGF)

SUMMARY

Scope of the company submission

The company submission (CS) presents evidence for the clinical effectiveness and cost effectiveness of cabozantinib (CABOMETYX®) for the first-line treatment of patients with untreated locally advanced or metastatic renal cell carcinoma (RCC). Cabozantinib is an orally administered tyrosine kinase (RTK) inhibitor. The drug inhibits vascular endothelial growth factor (VEGF) and hepatocyte growth factor receptor protein (MET), implicated in tumour growth and angiogenesis, pathologic bone remodelling, drug resistance, and metastatic progression of cancer. The recommended dose is 60 mg once daily, with lower dose adjustments recommend to manage adverse reactions. Treatment continues until disease progression or the occurrence of unacceptable toxicity.

The patient population in the CS is adults with untreated, intermediate or poor risk (International Metastatic RCC Database Consortium (IMDC) criteria), locally advanced or metastatic RCC. The CS reports a comparison of the effects of cabozantinib versus sunitinib and versus pazopanib as initial therapy for patients with poor or intermediate risk metastatic RCC.

Summary of submitted clinical effectiveness evidence

Systematic literature searches were performed to identify relevant clinical effectiveness studies. Searches identified one randomised controlled trial (RCT) of relevance to the appraisal, the CABOSUN trial. No direct trial evidence comparing cabozantinib versus pazopanib was identified.

CABOSUN was an investigator-led open-label, phase II RCT conducted by the Alliance for Clinical Trials in Oncology and conducted in 77 centres in the USA. It compared cabozantinib against sunitinib as first-line treatment. The trial included adult patients (≥18 years of age) with untreated clear cell metastatic RCC, Eastern Cooperative Oncology Group (ECOG) performance status of 0 to 2 and intermediate or poor risk per IMDC criteria. Patients received 60 mg of cabozantinib (n=79) orally once per day or 50 mg of sunitinib (n=78) orally once per day (sunitinib: 4 weeks on and 2 weeks off), with treatment cycles for both trial arms defined as 6 weeks. Although not designed as a registration trial, the trial was used to support the marketing authorisation for cabozantinib for this indication (anticipated date of approval: May 2018) based on what the CS describes as “encouraging findings”. The trial is a key source of evidence for the company’s cost-effectiveness analysis. Based on the

requirements for the marketing authorisation, the CS presents retrospective analysis of this trial using assessment of tumour response and progression by an independent radiology committee (IRC), and using US Food and Drug Administration (FDA)-recommended censoring rules.

The primary trial outcome measure was progression free survival (PFS). Secondary outcome measures included: overall survival (OS), objective response rate (ORR) and adverse effects (AE) of treatment. Patient cross-over between trial arms was not permitted during the trial, however, upon disease progression patients in both arms received subsequent systemic non-radiation anti-cancer treatments (cabozantinib group 57%; sunitinib group 58%). Health-related quality of life (HRQoL) was not measured in the trial (alternative sources of HRQoL utility estimates were used in the economic model).

Generally, baseline characteristics between the treatment arms were balanced apart from the proportion of patients with ≥2 metastatic sites (cabozantinib group 79%; sunitinib group 67%).

Outcome data from the CABOSUN trial were reported for different data cut-off points. The ERG presents data in this report for the latest time-point available for each outcome: PFS - September 2016; OS - January 2017 and an updated analysis July 2017; and tumour response - September 2016.

Results of the CABOSUN trial

PFS

- At a median follow-up of 25 months (September 2016 data cut-off), median PFS was 8.6 months (95% confidence interval (CI) 6.8, 14.0) for cabozantinib and 5.3 months (95% CI 3.0, 8.2) for sunitinib (p=0.0008), with a median difference of 3.3 months.
- The hazard ratio (HR), stratified by IMDC risk category and bone metastases, was 0.48 (95% CI 0.31, 0.74).
- The majority of events recorded were for documented disease progression (cabozantinib 51%, sunitinib 55%). PFS at 12 months (% event free) was 43.1 and 21.1 in the cabozantinib and sunitinib groups, respectively.

OS

- At a median follow-up of 28.9 months for OS (January 2017 data cut-off), the median OS was 30.3 months (95% CI 14.6, not estimable) in the cabozantinib arm versus

21.0 months (95% CI 16.3, 27.0) in the sunitinib arm, with a median difference of 9.3 months. The data were immature at this data cut-off, and there was a notable degree of censoring around the median estimates (censoring due to no event as of the cut-off date: cabozantinib 52%, sunitinib 42%). Hence, the data should be interpreted with caution.

- The HR, stratified by IMDC risk category and bone metastases, was 0.74 (95% CI 0.47, 1.14) p=0.1700.
- The percentage of patients event-free at 30 months was 50.7% for cabozantinib and 30.3% for sunitinib.
- Updated OS results at the 1st July 2017 data cut off: median OS for cabozantinib was 26.6 months (95% CI 14.6, not estimable) versus 21.2 months (95% CI 16.3, 27) for sunitinib. The HR was 0.80 (95% CI 0.53, 1.21) 2-sided p-value = 0.29. These data are also immature.

ORR

The ORR was 20% (95% CI 12.0%, 30.8%) in the cabozantinib arm, compared to 9% (95% CI 3.7%, 17.6%) in the sunitinib arm. The difference between groups in ORR was 11.3% (95% CI, 0.4 22.2%; p=0.0406). The ORR was classed as a 'confirmed partial response'. There were no confirmed complete responders in either study group.

Subgroups

There was a consistently favourable effect on PFS for cabozantinib compared with sunitinib in pre-defined subgroups (e.g. age, sex, race, baseline ECOG status, bone metastases). Confidence intervals were wide and included 1 for some the smaller subgroups. Subgroup results for OS also showed a favourable effect for cabozantinib compared with sunitinib, however, in most subgroups the confidence intervals included 1. Caution is advised given the observational nature of subgroup data and small sample sizes.

Adverse events

The majority of patients had at least one treatment-related adverse event regardless of treatment arm (95%-97%). Around half experienced a serious adverse event (49%-51%) and just over a third of all patents had a treatment-related serious adverse event (36%). Over half of all patients experienced a Grade 3 or 4 adverse event (60%-63%).

Discontinuations of study drug due to adverse events was also similar between study groups (21%-22%). Patients receiving cabozantinib had longer treatment exposure compared to those receiving sunitinib (median: 6.5 months versus 3.1 months, respectively) and dose

reductions were frequent with both treatments (46% and 35%, respectively), as were dose interruptions (73% and 71%, respectively).

The percentage of patients dying up to 30 days after last dose of study treatment was higher in the sunitinib group compared to the cabozantinib group (11% versus 5.1%, respectively), as was the case for death > 30 days after last dose of study treatment (49% versus 44%, respectively).

The most common adverse events (of any grade) in the cabozantinib treatment group were diarrhoea (72%), fatigue (62%), aspartate aminotransferase increased (60%), hypertension (56%), alanine aminotransferase increased (54%), decreased appetite (45%) and palmar-plantar erythrodysaesthesia syndrome (42%). In the sunitinib group common adverse events were fatigue (67%), platelet count decreased (58%), diarrhoea (49%), anaemia (44%) hypertension (38%), nausea (36%) and neutrophil count decreased (35%).

Indirect treatment comparison

The company conducted indirect treatment comparisons (ITCs) to compare cabozantinib against pazopanib given the lack of head-to-head evidence for these two treatments. The company's ITCs include two RCTs: CABOSUN (cabozantinib versus sunitinib) and COMPARZ (sunitinib versus pazopanib). The comparison with pazopanib is made through the common comparator sunitinib.

Due to the company's observation that proportional hazards do not hold for all survival outcomes in both trials the company used two Bayesian statistical ITC methods that do not assume proportionality in hazards:

- The parametric survival curve method by Ouwens et al provides survival estimates for a family of parametric distributions (Weibull, log-logistic, log-normal, Gompertz, exponential) and can extrapolate outcomes as described by two parameters (shape and scale);
- The fractional polynomial method by Jansen provides survival estimates for first order and second order models from a set of powers (five models for each order, 10 models in total). From these 10 models a best-fitting model was chosen by the company (second order P1=-1 and P2=-1) based on the deviance information criteria.

Both the Ouwens et al and fractional polynomial methods provide OS and PFS effect estimates that are used in the company's economic model.

The CS reports the results of the ITC as fitted survival curves for the outcomes of OS and PFS for all three treatments (cabozantinib, sunitinib, pazopanib), based on fixed effect and on random effects, for each of the five parametric distributions generated by the Ouwens et al method. For each of the analyses cabozantinib had a higher survival estimate than sunitinib or pazopanib. The sunitinib and pazopanib curves were similar to each other in shape and position, indicating similar effectiveness between these two treatments.

The CS presents fitted fractional polynomial survival curves for the outcomes of OS and PFS for all three treatments, based on fixed effects for first and second order models. On request the company also supplied HR plots with credible intervals for each fractional polynomial model to allow visual inspection of the time-varying HR curves. Results for PFS from the best-fitting fractional polynomial model (which informs the economic model base case) show:

- The HR for pazopanib peaks at month four [redacted] and declines slightly during the rest of the follow-up period. The HR for sunitinib peaks at month six [redacted] and declines slightly during the remainder of the follow-up period.
- The credible intervals increase over the follow-up period, with the upper bound increasing to include 1 after month 19 for pazopanib, and after month 11 for sunitinib.
- The time-varying PFS HRs for cabozantinib versus sunitinib generated by this fractional polynomial model compare broadly with the constant HR reported in the CABOSUN trial (0.48 (95% CI 0.31, 0.74)), though with greater uncertainty (wide credible intervals).

Results for OS from the best-fitting PFS fractional polynomial model (which informs the economic model base case) show:

- The HR for pazopanib starts to peak at month nine, and declines slightly after month 19 [redacted] The HR for sunitinib begins to plateau at month 13 and peaks at month 30 where it remains for the rest of the follow-up period [redacted].
- The credible intervals widen during the course of the follow-up period, and include 1 at all time points.
- The time-varying OS HRs for cabozantinib versus sunitinib generated by this fractional polynomial model compare broadly with the constant OS HR reported in the CABOSUN trial (0.80 (95% CI 0.53, 1.21), though with greater uncertainty (wide credible intervals).

Across the other fractional polynomial models (first and second order), the time-varying HR curves for cabozantinib versus sunitinib and cabozantinib versus pazopanib have a similar

shape to each other. Cabozantinib is of superior effectiveness when compared with both sunitinib and with pazopanib, with little difference between the results of each pairwise comparison.

The ERG considers that the statistical methods used to conduct the ITC are appropriate, but there is uncertainty in the results due to differences between the trials in patient prognostic characteristics (more detail on the critical appraisal of the ITC is available below under 'Commentary on the robustness of submitted evidence').

Summary of submitted cost effectiveness evidence

The CS includes:

- A review of published cost-effectiveness studies relating to cabozantinib, sunitinib and pazopanib in previously untreated locally advanced or metastatic RCC.
- An economic evaluation undertaken for the NICE STA process, comparing cabozantinib with pazopanib and sunitinib in treatment-naïve patients with locally advanced or metastatic RCC.

The company conducted a systematic search of the literature to identify economic evaluations with cabozantinib, sunitinib or pazopanib in untreated advanced RCC. The search identified 23 published cost-effectiveness studies, of which seven were conducted from an English, Welsh or British perspective. The company concluded that as none of the studies included cabozantinib, they are not directly relevant to this appraisal.

The company developed a model to evaluate the cost-effectiveness of cabozantinib as first-line treatment for advanced RCC. The model is a health state transition model, containing three mutually-exclusive health states: progression free (PF); progressed disease (PD) and death. Patients start in the PF state, and at disease progression, transition to the PD state, which is considered irreversible. Patients in PF and PD states die from cancer or other causes.

The distribution of the cohort between the health states and treatment states at each time point is estimated using a partitioned survival approach, based on PFS, Time to Treatment Discontinuation (TTD) and OS curves:

- **Death**: The proportion of patients alive at each time point is taken from the OS curve. Hence, the proportions of the cohort who die in each cycle are calculated.

- **PF**: The proportion of patients who are progression free is the minimum of the PFS curve and the OS curve at each time point.
- **PD**: The proportion of patients in the PD state is calculated as the residual (if any) of the cohort who are not dead and not progression free.

Patients enter the PF state on first-line treatment but may stop at any time due to adverse effects or when their disease progresses. After a fixed waiting period of 8 weeks, most patients then progress to subsequent treatment with one of 10 drugs included in the company's base case. The duration of second-line treatment is defined for each drug, after which patients are assumed to receive supportive care until death. The proportion of patients on first-line treatment is determined by the minimum of the TTD and PFS curves. Subsequent treatment status is calculated based on a waiting time and defined treatment duration for each individual second-line drug.

Other key features and assumptions of the model are listed below:

- **Cycle length**: 1 week, with half cycle correction.
- **Time horizon**: 20 years in base case (with 10 years in scenario analysis).
- **Duration of treatment effects**: based on extrapolation of PFS and OS curves fitted to trial data, assuming no waning of benefits over the time horizon.
- **Adverse events**: For each first-line treatment, grade 3 or 4 Treatment Emergent Adverse Events (TEAEs) with an incidence of 5% or more are included in the model. There is no explicit modelling of adverse events related to subsequent treatments.
- **Utility and QALY calculations**: Utility weights for the PF and PD health states are based on published estimates, assumed independent of treatment. Additional disutilities are applied to reflect included TEAEs for first-line treatments – applied as a one-off QALY loss in the first cycle. QALYs are also adjusted for the gender mix and age of the cohort.
- **Health resource use and costs**: The model estimates costs for of first-line and subsequent treatment; monitoring and disease management in PF and PD states; treatment of TEAEs for first-line treatments; and end of life care, applied in the last cycle before death.
- **Discounting**: 3.5% per year for costs and QALYs.
- **Uncertainty**: the model allows for exploration of uncertainty over input parameters using deterministic sensitivity analysis; scenario analyses varying selected model assumptions; and probabilistic sensitivity analysis (PSA) to estimate the joint effects of parameter uncertainty on the estimated costs and QALYs.

To apply the partitioned survival model, OS, PFS and TTD curves are required for cabozantinib and comparators, extrapolated over the 20-year time horizon. The company present two sets of base case results:

1. **Direct comparison** (cabozantinib versus sunitinib)
   This analysis is based on patient-level data from the CABOSUN trial, with OS, PFS and TTD curves separately fitted for cabozantinib and sunitinib arms using six families of survival functions: exponential, Weibull, Gompertz, log-normal, log-logistic and generalised gamma.
2. **Indirect comparison** (cabozantinib versus sunitinib and pazopanib)
   The two ITC meta-analyses methods described above were used to estimate PFS and OS curves:
   - ITC parametric curves, fixed and random effect models for five survival functions: exponential, Weibull, Gompertz, log-normal, log-logistic. The generalised gamma distribution was not implemented due to the lack of the incomplete gamma function in WinBUGS. The company reports that treatment was tested as a covariate, but the model only includes curves that were fitted separately for cabozantinib and sunitinib.
   - ITC fractional polynomial curves, with five first-order and five second order functions.

   As TTD data are not available from COMPARZ, the company uses curves fitted to CABOSUN for cabozantinib and sunitinib, and assumes that the latter also apply to pazopanib. This assumption is justified by the similar median and mean duration of treatment between the treatment arms in COMPARZ.

Survival curves for PFS, OS and TTD used in the company's base case and scenario analyses are summarised in the table below.

Table 1 Survival curves used in company analyses

Curve	Method	Treatment	CS Base case	CS scenarios
PFS	**Direct** CABOSUN	Cabozantinib Sunitinib	Log-normal	Exponential Weibull Gompertz
	ITC CABOSUN & COMPARZ	Cabozantinib Sunitinib Pazopanib	FP P1=P2=-1	RE exponential RE Weibull RE Gompertz
OS	**Direct** CABOSUN	Cabozantinib Sunitinib	Exponential	Exponential Weibull Gompertz
	ITC CABOSUN & COMPARZ	Cabozantinib Sunitinib Pazopanib	FP P1=P2=-1	RE exponential RE Weibull RE Gompertz FP P1=-0.5, P2=0 FP P1=-1, P2=0
TTD	**Direct** CABOSUN	Cabozantinib Sunitinib Pazopanib	Log-normal	Exponential Weibull Gompertz Generalised gamma

PFS, progression free survival; OS, overall survival; TTD, time to treatment discontinuation; ITC, indirect treatment comparison; RE, random effects; FP, fractional polynomial

Base case utility estimates for the PF and PD health states were taken from a previous NICE appraisal (Tivozanib TA512). The company also tested scenarios using health state utility estimates from the NICE appraisals for sunitinib and pazopanib, and the Swinburn et al 2010 study. These utility estimates are shown in the table below.

Table 2 Utility values (adapted from CS Tables 46 & 47)

Health state	Utility value: mean (SE)	95% CI	CS reference
Progression free			
Base case	0.726 (0.011)	0.705 to 0.748	Tivozanib TA512
Scenario	0.70 (0.01)	0.680; 0.720	Pazopanib TA215
Scenario	0.78 (0.078)*	0.627; 0.933*	Sunitinib TA169
Scenario	0.795 (0.0176)	0.761; 0.830	Swinburn 2010
Progressed disease			
Base case	0.649 (0.019)	0.612 to 0.686	Tivozanib TA512
Scenario	0.59 (0.059)*	0.474; 0.706*	Pazopanib TA215
Scenario	0.705 (0.071)*	0.567; 0.843*	Sunitinib TA169
Scenario	0.355 (0.0288)	0.299; 0.412	Swinburn 2010
TEAE grade ¾			
Base case	-0.2044 (0.0682)	-0.0707 to -0.3381	COMPARZ
Scenario	-0.0550 (0.0068)	-0.0418; -0.0685	METEOR trial

Abbreviations: TEAE, treatment emergent adverse effects; CI, confidence interval; SE, standard error; *SE or 95% CI not available in literature; 10% of the mean assumed.

Utility estimates for adverse events are sourced from the COMPARZ (base case analysis) and METEOR (scenario analysis) trials (METEOR was a phase III trial which compared cabozantinib with everolimus in patients with RCC that had progressed after VEGFR-targeted therapy[51]). The company assumes that the utility effects of adverse events are not disease-specific and that all types of grade 3 or 4 events elicit the same utility loss for a fixed period of 4 weeks and a fixed number of episodes per patient per TEAE. These assumptions yield a mean QALY loss of 0.0225 per TEAE in the base case (0.006 in the METEOR trial-based scenario). The company models the incidence of grade 3/4 TEAEs based on reported rates from the CABOSUN study for cabozantinib and sunitinib and from COMPARZ for pazopanib. Only events with a reported incidence of 5% or greater in at least one arm were included.

The company conducted a systematic literature review to identify published resource use and cost data relevant to the cost-effectiveness analysis. The costs included in the economic model are acquisition and administration of first-line and subsequent treatments, with adjustment for dose intensity and wastage when appropriate; monitoring and disease management in PF and PD states; treatment of TEAEs for first-line treatments; and end of life care. The CS reports PAS prices for first-line treatment but not second-line treatment. The company consulted UK-based clinical experts for the estimation costs accruing from of health state management resources, adverse event resource use and end-of-life care.

The results of the economic model are presented as incremental cost effectiveness ratios (ICERs).

- For the company base case, using the direct comparison from the CABOSUN trial, an ICER of £37,793 per QALY gained is reported for cabozantinib versus sunitinib.
- Based on their preferred ITC model, sunitinib is dominated by pazopanib and the ICER for cabozantinib compared with pazopanib is £48,451. The pairwise ICER for cabozantinib compared with sunitinib in this model is £31,538.
- Probabilistic results were similar.

The company conducted one-way deterministic sensitivity analyses and concluded that the key drivers to the cost-effectiveness results include drug costs and discount rates for QALYs and costs. Other parameters identified in the company's one-way sensitivity analysis include relative dose intensity and utilities associated with the progression-free state. The company's scenario analyses found cost-effectiveness results to be most sensitive to the choice of OS curve used in the model.

Commentary on the robustness of submitted evidence

Strengths

- The literature searches conducted by the company were considered by the ERG to be appropriate and sufficiently comprehensive to have identified all the relevant clinical effectiveness evidence. The company's systematic review methods were considered appropriate.
- The CABOSUN trial provided a direct comparison with sunitinib for PFS, OS, tumour response and adverse events. The ERG considers it to be well conducted overall, though there is a lack of detail on randomisation and concealment of allocation procedures to inform assessment of risk of selection bias. The open-label nature of the trial means the potential risk of performance and detection bias. However, the retrospective blinded IRC assessment of tumour response and progression conducted for the regulatory submission reduces the risk of detection bias for PFS and tumour response outcomes. This trial has some further limitations as described below.
- The Ouwens and fractional polynomial ITC methods appear to have been implemented adequately in accordance with the original methodological publications and the ERG considers that both are suitable for use for the indirect comparison of treatments in this appraisal. However, the results of both methods may be biased by the differences in RCC risk factors and other variables between the CABOSUN and COMPARZ trials (see below).
- The company's systematic review of cost effectiveness was of good methodological quality. The ERG agrees with the company's conclusion that none of the studies identified from the literature review included cabozantinib, and as such, they are not directly relevant to this appraisal.
- The two studies (CABOSUN and COMPARZ) used to estimate outcomes of PFS and OS provide the best available data sources, although the ERG does have concerns about the differences in patient population. The company conducted a range of ITC curve fitting methods (parametric and fractional polynomial methods) and used the resulting curves to make the indirect comparison from cabozantinib to pazopanib and to extrapolate beyond the trial follow up.
- The structure of the company's model reflects the nature of progression and clinical pathway for people with previously untreated locally advanced or metastatic renal cell

cancer. The company used methods for the economic evaluation that are consistent with NICE methodological guidelines and with other drug appraisals for this population.

- The ERG agrees that the health state utility values applied in the company's model meet the NICE reference case and are suitable for inclusion in the model. Costing methods and sources are also generally of good standard with reasonable assumptions. Scenario analysis reflective of the current NHS practice are explored.

Weaknesses and areas of uncertainty

- The CABOSUN trial has some limitations.
    - It is a phase II trial, and was never designed to be a registration trial. It has a relatively small sample size (n=157 patients).
    - The trial was conducted entirely in the US and therefore it may not necessarily be applicable to the UK (though clinical experts to the ERG regarded the baseline characteristics as generally representative of patients in their practice).
    - OS was a secondary outcome and the data are immature. HRQoL was not an outcome measure.
    - The updated PFS assessment conducted for the regulatory submission (and used in the CS) used different censoring rules and a blinded IRC, which meant that the number of events (progressions or deaths) recorded (n=92) was less than the number required in the original PFS statistical power calculation (n=123). This means the updated PFS assessment would be statistically under-powered.
    - There was an imbalance between trial arms in the number of patients with missing data. One patient in the cabozantinib arm and six in the sunitinib arm withdrew prior to receiving study treatment, but the reasons for these withdrawals were not known. There was also a higher incidence of missing or unevaluable data in the sunitinib arm, with six patients in the cabozantinib arm and 18 in the sunitinib arm not evaluable because they had no adequate post-baseline imaging assessments. The CS states that based on their baseline characteristics (data unavailable to the ERG to verify), the sunitinib patients without post-baseline imaging would not be expected to have a better prognosis than sunitinib patients who had a response recorded, and therefore it is unlikely that the radiographic endpoints were biased against sunitinib by these missing data.
- There are some important differences between the two trials in the ITC:

    - The CABOSUN trial included only patients at intermediate or poor RCC risk, whilst the COMPARZ trial included patients at favourable, intermediate and poor risk classifications.
    - A greater proportion of COMPARZ patients were classified as having the highest cancer performance status, likely due to inclusion of some patients with favourable RCC risk status in the trial.
    - Around a third of patients in CABOSUN had bone metastases (a key prognostic factor in RCC) at baseline compared to 18% of patients in COMPARZ.
    - The impact of these differences on the results of the ITC are not discussed in the CS. The ERG considers that the impact of the differences on the ITC results to be uncertain.
- The ERG believes that the company's cost effectiveness results include some errors in model inputs and calculations, which could bias conclusions on cost-effectiveness. The ERG corrected errors in the company's QALY calculations and small errors in costs.
- It is appropriate to estimate costs and health effects over the patients' whole lifetimes, so we do not disagree per se with the company's use of a 20-year time horizon. However, other RCC NICE appraisals have adopted a more conservative time horizon of 10 years. In the company's base case model, a relatively small proportion of the modelled cohort survive to 10 or 20 years. However, we question the extrapolation of OS and PFS curves from limited trial follow-up over 20 years. This entails strong assumptions about persistence of treatment effects, which may not be realistic. We investigate the impact of the time horizon and different assumptions about waning of treatment effects in ERG analysis.
- Although the company's preferred survival models have reasonable face validity with good measures of fit, they appear to overestimate PFS and OS. We note that other fitted models do not necessarily address this uncertainty. Based on measures of fit and plausibility of extrapolation, the ERG agrees with the company's selection of best direct comparison and ITC parametric and fractional polynomial curves. However, selection of curves for scenario analyses fit less well. We explored alternative assumptions in the ERG scenario analysis. For the direct comparison, we note that the company's choice of the exponential distribution for both cabozantinib and sunitinib conflicts with the conclusion that OS hazards are not proportional. However, we suggest that the exact shape of the CABOSUN Kaplan-Meier (KM) OS curves should not be over-interpreted given the modest sample size (n=157) and lack of explanation

for why the curves should come together and then diverge between about 13 and 20 months.

- Median survival for OS and hazard ratio estimates are less favourable for the most recent data cut-off (July 2017) than in the earlier cut-off of January 2017 used to fit OS in the model (CS B.2.6 Figures 6 and 7). (NB. The CS does not explicitly state which OS dataset was used to inform in the model, but the January 2017 KM plot is reproduced in the CS economic chapter and KM data provided by the company in response to a clarification question also relates to this earlier cut-off). This suggests that the model may over-estimate the survival advantage for cabozantinib over sunitinib.
- The ERG considers that it is highly unlikely that the QALY loss is the same for all types of TEAE, but that these assumptions reflect a reasonable average. We conduct additional scenario analysis to test model sensitivity to the TEAE disutility parameter, including higher as well as lower estimates of the disutility. In addition, we note that of 59 types of adverse events listed in the company's model, only 18 events with incidences equal to or greater than 5% were modelled. We test the impact of changing the inclusion threshold for TEAEs in scenario analysis.
- The model does not include an adjustment for age-related increase in mortality in the general population, as the model relies entirely on the projected OS curves. However, given the high rate of mortality for people with advanced RCC, this might not affect results. We check that the model does not yield counter-intuitive results with longer-surviving RCC patients having lower mortality than members of the general population at the same age.

Summary of additional work undertaken by the ERG

We corrected the company's model to reflect the identified errors. The most significant were coding errors in QALY calculations that had the effect of underestimating QALYs for each treatment, and hence underestimating the incremental QALY gain with cabozantinib compared with sunitinib and pazopanib. There were also small discrepancies in some cost estimates. The corrected model resulted in lower ICER estimates for the company's base case:

- £31,956 per QALY for the direct comparison of cabozantinib with sunitinib;
- £40,757 for cabozantinib compared with pazopanib and £26,182 compared with sunitinib based on the ITC analysis.

These estimates are subject to uncertainty, with the method of fitting the OS curves and choice of survival function having the largest impact on the ICERs.

Probabilistic analysis estimated a 28% probability of the ICER compared with pazopanib being less than £30,000 per QALY gained in the ITC base case.

We conducted additional analyses to test alternative assumptions and scenarios. The ERG-preferred set of assumptions included the following key differences from the company base cases:

- **Method of fitting OS curves**. Due to our concerns about the robustness of the ITC, we prefer to rely on the analysis of CABOSUN data for direct comparison of cabozantinib with sunitinib. Although the proportional hazards assumption appears not to hold, we agree with the company that the exponential distribution gives the best balance of fit to the trial data for both treatment arms and plausible long-term extrapolations. We base the OS curve for sunitinib on the exponential curve fitted to CABOSUN data. We then estimate the cabozantinib OS curve using the reported hazard ratio from the most recent update of trial data (July 2017 data cut) – the company's analysis uses an earlier dataset (January 2017). Finally, we assume equivalent OS for pazopanib and sunitinib, based on the results of COMPARZ.
- **PFS and TTD curves**. We follow the company's direct base case for estimates of PFS and TTD for cabozantinib and sunitinib: with lognormal curves separately fitted by treatment to CABOSUN data. For pazopanib, we again assumed equivalence with sunitinib for time to progression based on the results of the COMPARZ trial.
- **Time horizon and duration of effects**. The company uses a 20 year time horizon, which is longer than in other recent appraisals for RCC. We believe that it is correct to reflect a whole life time horizon, so also use 20 years in our base case. However, we do not believe that it is appropriate to assume persistence of treatment effects for cabozantinib based on the limited trial follow-up and sample size. The ERG therefore adopts a conservative assumption that progression and mortality hazards for cabozantinib equal those of sunitinib after a fixed period of time: 5 years from baseline in our preferred analysis.
- **Health state utilities, adverse effects and costs**. The company approach to modelling the utility and cost impacts of the treatments were generally reasonable and reflected the NICE base case and decisions in previous appraisals. We therefore adopt the same base case parameters, but conduct some additional scenario analyses to test the robustness of the results.

The ERG preferred analysis gave estimated ICERs of £65,742 for cabozantinib compared with pazopanib and £41,465 compared with sunitinib (Table 3). As in the company base

case, we estimate that sunitinib is dominated by pazopanib due to its higher cost and similar effectiveness.

Table 3 Cost-effectiveness: ERG preferred assumptions analysis results

Drug	Costs (£)	QALYs	Life-years	PF life years	ICER (£ per QALY gained)	
					Incremental analysis	Pairwise, cabozantinib vs. comparator
Pazopanib	[redacted]	[redacted]	[redacted]	[redacted]	-	65,743
Sunitinib	[redacted]	[redacted]	[redacted]	[redacted]	-	41,465
Cabozantinib	[redacted]	[redacted]	[redacted]	[redacted]	65,743	-

However, this result was sensitive to some cost and resource use assumptions. By assumption, our preferred analysis gave the same life expectancy with sunitinib as with pazopanib, yielding very similar QALY estimates. Cabozantinib has a modest survival advantage and a larger effect on progression free survival and hence QALYs. We believe that these results appropriately reflect evidence from CABOSUN and COMPARZ. The results were generally robust, with the ICERs remaining above £30,000 per QALY gained for all of the scenarios that we tested.

The above analyses include existing PAS discounts for cabozantinib, sunitinib and pazopanib for first-line treatments. However, they exclude these arrangements and other existing PAS discounts for subsequent treatment after failure of first line treatment. We present results for the ERG-corrected company base case and scenarios and for ERG additional analysis in a confidential addendum to this report.

The ERG is of the opinion that cabozantinib does not fully meet the NICE criteria for being considered as a life-extending treatment for people with a short life expectancy. This is because the submitted CS model and results from the ERG's preferred assumptions give mean OS estimates exceeding 24 months for sunitinib and pazopanib.

1 Introduction to ERG Report

This report is a critique of the company's submission (CS) to NICE from Ipsen Ltd UK on the clinical effectiveness and cost effectiveness of cabozantinib for untreated locally advanced or metastatic renal cell carcinoma (RCC). It identifies the strengths and weakness of the CS. Clinical experts were consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the CS was requested from the company by NICE and the ERG on 22nd February 2018. A response from the company via NICE was received by the ERG on 9th March 2018 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The ERG considers that generally the CS provides a clear and accurate overview of the nature and clinical consequences of renal cell carcinoma (RCC).

2.1.1 Renal cell carcinoma (RCC)

RCC is a cancer that usually originates in the lining of the proximal renal tubules of the kidney - the smallest tubes inside the nephrons that help filter the blood and make urine. As stated in the CS, this is the most common type of kidney cancer, accounting to around 80% of all kidney cancer cases.[1] The three main types of RCC are clear cell (75%), papillary (10%) and chromophobe (5%).[2]

As stated in the CS, early small RCC tumours are usually asymptomatic and are often discovered incidentally during other investigations.[3] [4] In consequence, many patients present with advanced disease (around 38%[5]) and 25-31% of patients present with metastases at diagnosis.[2] The NICE final scope covers both locally advanced RCC (which cannot be removed by surgery) and metastatic RCC.

2.1.2 Clinical presentation

The most common symptom of RCC is blood in the urine (in around 50% of cases).[2] As described in the CS, other non-specific symptoms include weight loss, fever, sweating, fatigue and anaemia amongst others. At the metastatic stage, the tumour has spread beyond the regional lymph nodes to other parts of the body. Less frequent symptoms related to the metastatic spread of the disease include bone pain, skeletal-related events and

hypercalcaemia, as well as venous thromboembolism and lung symptoms such as airway obstruction.[3 6]

2.1.2.1 Staging and prognosis

A staging system is used to show how far the cancer may have spread (and whether it has spread into nearby lymph nodes or distant organs) on a scale of I to IV. Lower stage cancers are less likely to spread than higher stages cancers.[2] The NICE scope denotes stage IV (metastatic) cancer.

One of the most common staging systems (the extent of the cancer in the body) used is the American Joint Cancer Committee (AJCC) Tumour Node Metastasis (TNM) system. This classifies the size of the tumour. In addition to this, the CS presents the International Metastatic RCC Database Consortium (IMDC) risk stratification model (also known as 'the Heng model'). This is the method specified in the NICE final scope and cited in the Summary of Product Characteristics (SmPC). The IMDC is an update of a previous classification system known as the Memorial Sloan Kettering Cancer Center (MSKCC) model.[7 8] The MSKCC model is similar to the IMDC criteria, with the latter was a minor revision of the former.

According to expert clinical advice received by the ERG, the IMDC classification is not formally used in clinical practice in the UK. Level of risk has traditionally been judged on general clinical assessment and blood tests. As newer drug therapies are introduced targeted at specific risk groups use of the IMDC will likely increase. Clinical expert advice to the ERG indicates that the use of the IMDC classification would not require any significant changes to clinical practice. In this method, patients are assessed for the presence of six risk factors (routinely collected in practice):

- Karnofsky performance status (PS) <80%
- Haemoglobin <lower limit of normal
- Time from diagnosis to treatment of <1 year
- Corrected calcium above the upper limit of normal (ULN)
- Platelets greater than the upper limit of normal
- Neutrophils greater than the upper limit of normal

Based on the six risk factors, patients are categorised into three risk groups, which predict survival and influence the management of the patient's RCC:[7]

- Favourable – 0 factors
- Intermediate - 1 or 2 factors
- Poor - >3 factors.

The IMDC model has been externally validated in patients with metastatic RCC who were treated with first-line VEGF-targeted treatment, including patient stratification by risk (favourable risk group median overall survival 43.2 months after the start of targeted treatment, intermediate risk group 22.5 months and poor risk group 7.8 months).[7] The CS states that around 80% of all metastatic RCC patients are in the latter two risk groups and clinical experts advising the ERG concur with this. The CS cites a 5-year relative survival rate for stage IV RCC (i.e. metastatic) by Cancer Research UK as around 6% in the UK.[2]

2.1.3 Effects of RCC on health-related quality of life

The top five symptoms reported in a national, cross-sectional study by patients with advanced metastatic RCC are: fatigue, weakness, worry, shortness of breath, and irritability.[9] HRQoL in this patient group is also impaired by disease-related factors associated with tumour burden, for example anorexia-cachexia syndrome (associated with weight loss, lethargy, as well as possible fever, night sweats and distortion of the sense of taste amongst others), hypercalcemia, venous thromboembolism, pain (somatic, visceral and neuropathic), and metastases-associated specific site symptoms.[10]

Patients with advanced RCC generally have a poor prognosis and this, combined with the symptoms associated with advanced disease, can significantly affect all domains of patients' HRQoL not just physical functioning, such as emotional and social wellbeing and.[10 11] As might be expected, evidence shows that the effects of disease progression in these patients is linked to a deterioration in HRQoL.[12 13 14 15]

2.1.4 Epidemiology

The company provides an overview of the incidence of kidney cancer in the UK, mostly based on data reported by Cancer Research UK and the National Office of Statistics. Figures of new cases of kidney cancer for England in the CS are cited for 2015, with 9023 new cases (ICD-10 C64 malignant neoplasm of kidney, except renal pelvis), equating to an age-standardised rate of 24.3 per 100,000 in males and 12.3 per 100,000 in females. More recent data identified by the ERG by the Office for National Statistics in England shows that during 2016, 5823 new cases of kidney cancer for males and 3392 for females were recorded (an increase of over 2%), equating to age-standardised rates of 24.5 per 100,000 in males and 12.4 per 100,000 in females.[16] RCC is a sub-type of kidney cancer, accounting for around 80% of all kidney cancer cases, as stated above.

Kidney cancer is the UK's seventh most common cancer, accounting for 3% of incident cases. Partly due to increased detection of early-stage tumours, rates in the UK are estimated to increase annually by 1.2%.[2] There were a reported 3319 deaths from kidney cancer in 2015 in England, with no updated figures as yet available for 2016.[16]

Kidney cancer is more predominant in males, described as a 17:10 male:female ratio in the CS. The ERG note that this was reported as 63% vs 37% (male:female) in the CS-cited source.[2] In the UK, it is the sixth most common cancer in men and the 10th in women and incidence increases with age.[2] For both men and women, the highest rates of kidney cancer are in the 85 to 89 age group.[2]

The risk factors reported in the CS are for kidney cancer only and not specific to RCC: 42% major lifestyle and other risk factors, 62% for hypertension, 24% for smoking and 24% for excess bodyweight. However, cigarette smoking, obesity and hypertension are well-established risk factors for RCC.[17] The risk factor of related hereditary syndromes is not reported, most likely because its occurrence is relatively low (approximately 3% to 5%).[18]

2.2 Critique of company's overview of current service provision

The CS provides a generally clear and accurate overview of how locally advanced and metastatic RCC is managed in clinical practice.

Advanced RCC is incurable and largely resistant to chemotherapy, radiotherapy and hormonal therapy. Due to the lack of improved survival with either chemotherapy or hormonal therapy alone, the mainstay of treatment for locally advanced or metastatic RCC starting in the late 1980s were cytokines, of which interferon alfa and interleukin-2 have been the most evaluated.[19] Targeted drug therapies are now the mainstay of treatment, although some patients receive surgery to reduce the size of the tumour or to remove metastases and this may be in addition to drug treatment.[1] The CS states that treatment goals are to extend life, delay disease progression, relieve symptoms and maintain function, citing a previous NICE appraisal (TA178) as reference.[20] Figure 1 illustrates the NICE pathway of care for renal cancer.[21]

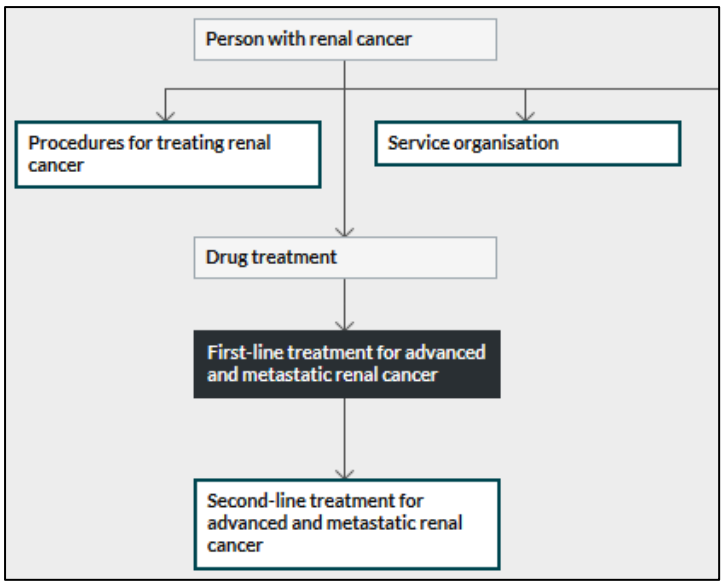


Figure 1 NICE pathway of care in renal cancer

Currently recommended first-line treatments for previously untreated advanced RCC by NICE are:

- Sunitinib in patients who are suitable for treatment and have an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 (TA169).[12]
- Pazopanib, in patients who have not received prior cytokine therapy and have an ECOG performance status of 0 or 1 (TA215).[13]
- Tivozanib for treating advanced renal cell carcinoma in previously untreated adults (TA512).[22] (NB. This drug was not included in the scope of the current NICE appraisal as its appraisal had not completed at that time).

Bevacizumab, sorafenib and temsirolimus are not recommended by NICE for people with advanced and/or metastatic RCC (TA178).[20]

As there are no UK-specific clinical guidelines for the treatment of RCC, the CS states that in addition to the medicines recommended by NICE, current clinical practice in England and Wales reflects the following guidelines:

- European Society of Medical Oncology (ESMO) Renal Cell Carcinoma: Clinical Practice Guidelines for diagnosis, treatment and follow-up.[1]
- Updated European Association of Urology (EAU) Guidelines: Recommendations for the Treatment of First-line Metastatic Clear Cell Renal Cancer.[23]
- National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology, kidney cancer.[6]

2.2.1 Proposed place of cabozantinib in the clinical pathway

The CS states that it is anticipated that cabozantinib (in this indication) will be used in accordance with its marketing authorisation ("treatment of advanced renal cell carcinoma (RCC) in treatment-naïve adults with intermediate or poor risk per IMDC criteria"). The CS proposes that cabozantinib would be an additional treatment option alongside sunitinib and pazopanib in intermediate or poor risk patient groups.

Both the EAU and NCCN guidelines have been updated to include cabozantinib as a treatment option in previously untreated IMDC intermediate and poor risk RCC, while the position of the ESMO guidance on cabozantinib as a treatment option in previously untreated RCC is still unclear.

2.2.2 Potential impact on current service provision

As cabozantinib is another orally administered treatment, the CS states that there is no requirement for a change in current management arrangements or infrastructure. The CS states that testing required to assign patients to IMDC risk groups is carried out as part of routine clinical practice. As stated above, the IMDC criteria are not formally used in clinical practice in the UK, but clinical advice to the ERG is the information required to complete the criteria are routinely collected.

Treatment dose modifications can be managed remotely, without the patient having to attend a consultation in person. Cabozantinib is already recommended by NICE within its marketing authorisation for use in previously treated advanced RCC (TA463[14]) and the ERG therefore agrees that there should be no additional impact on current service provision when used in a previously untreated patient group.

2.3 Critique of company's definition of decision problem

The CS provides a summary table (CS Table 1) including the final decision problem issued by NICE, the company's decision problem and a rationale for any differences between the two.

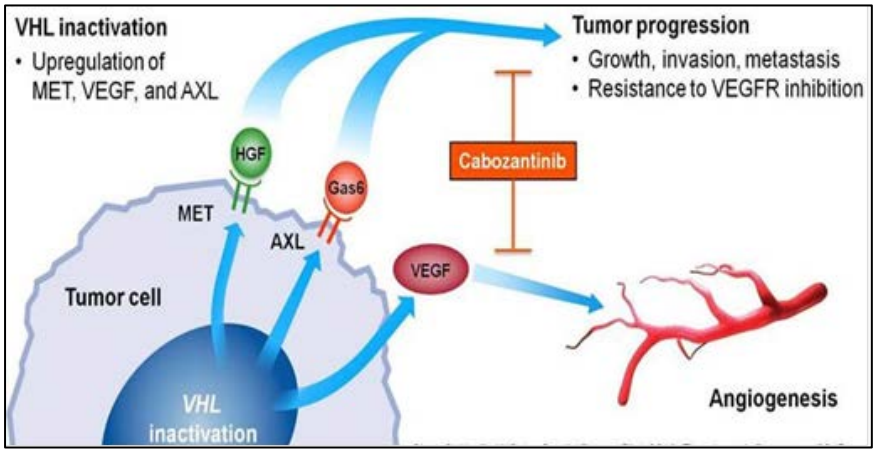
2.3.1 Population

The population specified in the company's decision problem is people with untreated, intermediate or poor risk (as per IMDC criteria), locally advanced or metastatic RCC. The CABOSUN trial (which is the main cabozantinib clinical effectiveness study in the CS) focused on IMDC intermediate- and poor-risk groups, the rationale being that these groups capture 70% to 80% of all patients with advanced disease and because such patients are most in need of systemic therapy and disease control.[24] The patient population matches that

specified in the final scope issued by NICE and that specified in the SmPC indication for cabozantinib (application for marketing authorisation for cabozantinib for "the treatment of advanced RCC in treatment-naïve adults with intermediate or poor risk per IMDC criteria" was submitted to the European Medicines Agency (EMA) on 28 August 2017. Cabozantinib is already licensed for the treatment of advanced RCC in adults following prior VEGF-target therapy.

2.3.2 Intervention

In accordance with the final NICE scope, the intervention described in the company's decision problem is cabozantinib (brand name CABOMETYX®). Cabozantinib is a small molecule that inhibits multiple receptor tyrosine kinases (RTKs) implicated in tumour growth and angiogenesis, pathologic bone remodelling, drug resistance, and metastatic progression of cancer. Figure 2 shows that cabozantinib has a multi-targeted mechanism of action in the treatment of RCC, targeting and inhibiting the MET (hepatocyte growth factor receptor protein), VEGF (vascular endothelial growth factor) and AXL receptors.



Source: CS Section B.1.2 Figure 1

Figure 2 Cabozantinib's mechanism of action

The company supplied the SmPC with their submission to NICE. The revised European Public Assessment Report (EPAR) is not yet available. On March 22nd 2018 the Committee for Medicinal Products for Human Use (CHMP) expressed a positive opinion on the use of cabozantinib for first line treatment of adults with intermediate or poor risk advanced RCC. As outlined in the CS, the SmPC states the recommended dose of cabozantinib is a once-a-day tablet of 60 mg, but is also available as 20 and 40 mg. Treatment should continue until the patient is no longer clinically benefitting from therapy (assessed as tumour progression) or until unacceptable toxicity occurs. Suspected adverse drug reactions may require temporary treatment interruption and/or dose reduction. When dose reduction is necessary, it is recommended to reduce cabozantinib to 40 mg daily and then to 20 mg daily. Dose

interruptions are recommended for management of Common Terminology Criteria for Adverse Events (CTCAE) grade ≥3 toxicities or intolerable grade 2 toxicities. Dose reductions are recommended for events that, if persistent, could become serious or intolerable. Cabozantinib, the intervention described in the decision problem, is appropriate for the NHS and reflects its licensed indication.

2.3.3 Comparators

The two comparators of interest listed in the company's decision problem are those specified in the NICE final scope:

- sunitinib
- pazopanib

These comparators are appropriate for the NHS as they have been recommended for first line use by NICE. As previously stated, both sunitinib[12] and pazopanib[13] are licensed as first-line treatment of advanced and/or metastatic RCC.

2.3.4 Outcomes

The company has listed all but one of the outcomes specified in the NICE final scope in their decision problem:

- overall survival (OS)
- progression-free survival (PFS)
- response rates
- adverse effects (AE) of treatment

The NICE finial scope specified health-related quality of life (HRQoL) as an outcome, but no such data were collected in the single phase II trial (CABOSUN) presented in evidence of the clinical effectiveness of cabozantinib. Hence, HRQoL was not presented as a clinical effectiveness outcome measure in the CS (though HRQoL utility data from other sources are used in the economic model).

2.3.5 Economic analysis

The partitioned survival model used in the CS is considered as one of the standard methods for population-based cancer survival analysis and the method is in line with previous health economic analyses.[13 14] (see section 4 of this report for description and critique of the company's economic evaluation).

2.3.6 Other relevant factors

There were no subgroups of relevance noted in the NICE final scope or the company's scope. Although the scope does not require subgroups to be assessed, the CABOSUN trial included subgroup analyses based on a number of factors, including RCC risk and bone metastases which are of prognostic significance.

The company states that they do not anticipate that the use of cabozantinib will be associated with any equality issues.

3 CLINICAL EFFECTIVENESS

3.1 Critique of company's approach to systematic review

3.1.1 Description of company's search strategy

The CS reports four separate systematic literature searches:

- Clinical effectiveness evidence. The search strategy used in the submission to NICE by the manufacturer of pazopanib in NICE TA215.[13] covering the period 1980 to 2009 was adapted by the company and updated to 28th June 2017 (search strategy reported in CS Appendix D).
- Cost effectiveness evidence. Search period: 1946 to 19th September 2017 (search strategy reported in CS Appendix G).
- Health Related Quality of Life (HRQoL). Original search period: 2006 to July 2016; Update search period: 2016 to 28th July 2017 (search strategy reported in CS Appendix H).
- Cost and healthcare resource identification measurement and valuation. Original search period: 2006-2016; Update search: 2016 to 19th September 2017 (search strategy reported in CS Appendix I).

The clinical effectiveness search strategy was designed from a global perspective. It included search terms for a range of treatments including those within the scope of the appraisal (cabozantinib, pazopanib and sunitinib) and others not in the scope (interferon alfa, interleukin-2, sorafenib, bevacizumab and interferon alfa, temsirolimus, tivozanib). The search terms contain appropriate subject headings together with a good range of truncated free text. An appropriate range of databases was searched: Medline (including In-Process and other non-indexed citations); Embase and the Cochrane Library. A combined search filter was used to identify RCTs, controlled and other trials, meta-analyses and systematic

reviews. The clinical effectiveness search strategy is extensive but visually overcomplicated with various date restrictions applied to different sets of drugs. The ERG notes, however, that the sets are correctly combined and the number of hits (records retrieved) per line is documented for transparency. The search write up offers guidance to the strategy, is thorough and transparent.

Supplementary searching was undertaken in the CS to identify ongoing trials on the National Institute of Health's (NIH) clinical trial registry (www.clinicaltrials.gov) and reference lists from other HTA submission documents were searched. Conference abstracts were not specified as being searched separately. The company state that there are no relevant ongoing studies (CS Section B.2.11), and in response to a clarification request (question A13) stated that they are not aware of any planned or ongoing trials of cabozantinib (as a single therapy agent) for the indication in this appraisal.

The ERG re-ran the company's clinical effectiveness searches on Medline, Embase and the Cochrane Library for the years 2017 to present, to identify any recently published relevant studies. The ERG additionally ran searches of two databases on the Web of Science Platform: Conference Proceedings Citation Index-Science (CPCI-S) and the Emerging Sources Citation Index (ESCI). The following conferences for the years 2016-2018 were additionally searched on the internet: American Society of Clinical Oncology (ASCO), European Society for Medical Oncology (ESMO), European Cancer Organisation (ECCO), European Multidisciplinary Meeting on Urological Cancers (EMUC), International Kidney Cancer Symposium (IKCS). Given that the CS only searched one on-going trials database the ERG checked for any missing ongoing trials on NIHR UKCTG (UK Clinical Trials Gateway), the WHO ICTRP (International Clinical Trials Platform) and re-checked the clinicaltrials.gov database.

Results from these searches were screened by an ERG reviewer. Two relevant conference abstracts not included in the CS were identified, both of which report results from the CABOSUN trial, the sole RCT of cabozantinib included in the company's systematic review of clinical effectiveness (see section 3.1.3 of this report). One of the abstracts[25] was linked to a poster which was included in the CS,[24] albeit the abstract contained less information than the poster. The second abstract[26] identified by the ERG was presented at the European Society for Medical Oncology (EMSO) conference in February 2018 and hence not available at the time the CS was produced. This abstract was linked to a slide presentation and a poster, both of which the ERG were unable to access. The abstract included a small amount

of additional information on tumour response results in a patient subgroup not presented in the CS. We report these data in section 3.3.6 of this report.

The cost effectiveness, HRQoL and health care resource-use searches are much easier to follow as the sets are grouped together more logically, without the varying date ranges. The terms and search filters are all appropriate.

In summary all searches are well documented and are fit for purpose and it is unlikely that any potentially relevant studies comparing cabozantinib with sunitinib and pazopanib were not included.

3.1.2 Statement of the inclusion/exclusion criteria used in the study selection

The company provides a description of the inclusion criteria for the systematic literature review (SLR) (CS B.2.1, Table 4), which was also used to identify studies for potential inclusion in an indirect treatment comparison (ITC) (see Section 3.1.7). These criteria were broader than the NICE final scope but the treatments were subsequently limited to cabozantinib, sunitinib and pazopanib for this appraisal. Details of the process and methods used to identify and select the clinical evidence relevant to the technology being appraised are contained in CS appendix D. The interventions and comparators reflect the nature of the decision problem, the anticipated licensed indication and current NHS practice. The CS provides a flow diagram illustrating the number of records identified through each of search sources: the electronic database search, the pazopanib company submission to NICE for TA215[13] and through study registry searches (clinicaltrials.gov) (CS Figure 3).

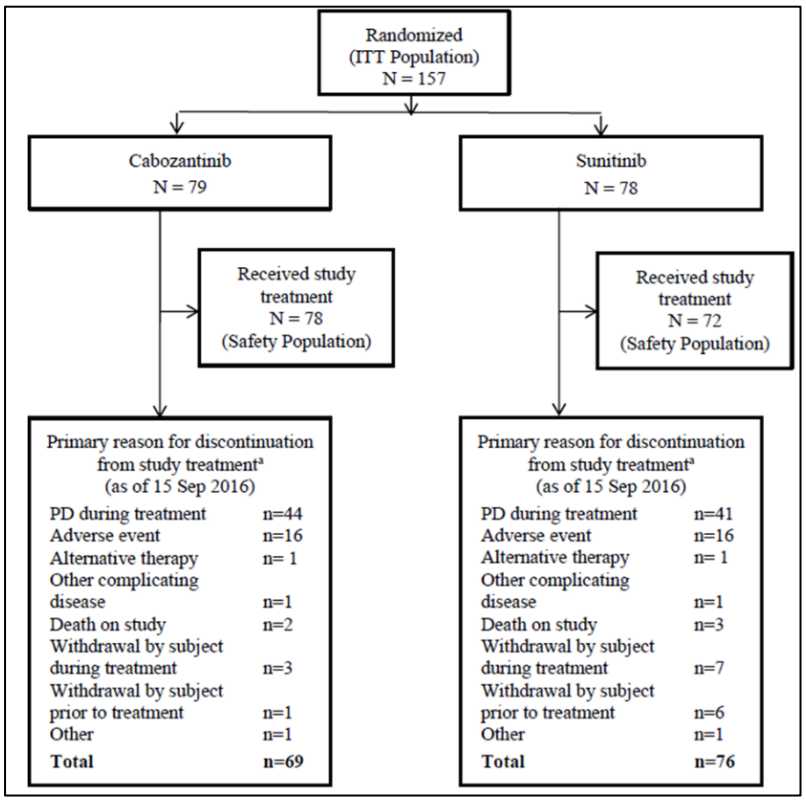
Reasons for the exclusion of studies at the full paper stage are provided (CS Figure 3) and references to these studies are listed in Appendix D1.1. Nine references are listed as 'article not obtained' in the flowchart and in response to a clarification request (question A12), the company states that these references were identified from the systematic review in the pazopanib company submission to NICE[13] (eight out of these nine references were conference abstracts). The company states that these references would either have been excluded or had been superseded by a more recent full text publication. It is our view that the non-availability of these nine references would not have biased the company's systematic review.

3.1.3 Identified studies

The CS identified one relevant published RCT, the A031203 CABOSUN trial (NCT01835158) referred to as CABOSUN for short. CABOSUN was an investigator-led phase II, open-label trial set in 77 centres in the USA from July 2013 to April 2015,

conducted by the Alliance for Clinical Trials in Oncology. The trial included adult patients (≥18 years of age) with untreated clear cell metastatic RCC, ECOG performance status of 0 to 2 and intermediate or poor risk per IMDC criteria comparing a cabozantinib treatment arm with a sunitinib treatment arm. The trial was supported by grants from the National Institutes of Health and by Exelixis (the manufacturer of cabozantinib, who provided the drug).

A CONSORT flowchart of the trial is presented in CS appendix D1.2, detailing the number of patients that discontinued/dropped out and associated reasons (see Figure 3).



Source CS appendix D.1.2 Figure 52

Figure 3 CABOSUN trial participant flow chart

Clinical effectiveness evidence is presented from the company study report (CSR)[27] and three journal publications, of which two were conference presentations.[24 28 29] The trial was used in support of the company's application for marketing authorisation, although not designed as a registration trial but used as such due to what the CS describes as "encouraging results". Due to requirements of the marketing authorisation, there are some discrepancies between the results presented in the CS and those in the trial journal publications (discussed in more detail in Section 3.1.6 of this report).

As can be seen in Table 4, the CABOSUN trial did not include any UK patients. Patients received 60 mg of cabozantinib orally once per day or 50 mg of sunitinib orally once per day (sunitinib: 4 weeks on and 2 weeks off), with a treatment cycle defined as 6 weeks for both. All patients regardless of treatment arm received full supportive care and AEs were managed through dose interruptions and dose reductions in both treatment arms.

Table 4 CABOSUN trial characteristics

<table>
<tr><th>Design, patient population and length of follow-up</th><th>Intervention</th><th>Comparator</th></tr>
<tr><td rowspan="3">Trial name: CABOSUN

Design: Phase II, open-label, multicentre RCT

Location: 77 centres in the USA

Setting: hospital and outpatient clinics

Number of participants: 157

Inclusion: Adults ≥ 18 years of age with documented RCC with some component of clear cell histology, that was advanced (defined as not amenable to curative surgery or radiation therapy) or metastatic (American Joint Committee on Cancer stage IV). Other key eligibility criteria were;
• Intermediate or poor risk by IMDC criteria
• ECOG performance status 0 to 2
• No prior systemic treatment for RCC
• No active brain metastases; patients with treated brain metastases which had been stable for at least 3 months were eligible
• Adequate organ and marrow function with no uncontrolled significant illness.

Length of follow-up:
(Randomisation July 2013 to April 2015) The median follow-up of surviving patients as of 15/09/2016 was 21.4 months.</td><td>Cabozantinib (n= 79) administered orally once per day at a dose of 60 mg</td><td>Sunitinib (n= 78) administered orally once per day at a dose of 50 mg for 4 weeks, followed by a 2-week break</td></tr>
<tr><td colspan="2">Adverse events were managed with treatment interruptions and dose reductions: cabozantinib to 40 and 20 mg, and sunitinib to 37.5 and 25 mg.

A treatment cycle was defined as 6 weeks in both study groups. Treatment duration was until disease progression, intolerance to therapy, or withdrawal of consent for treatment.</td></tr>
<tr><td colspan="2">Background therapy: all received full supportive care (including transfusions of blood and blood products, erythropoietin, antibiotics, antiemetics, and other agents) when appropriate.

Prophylactic measures were taken to prevent or reduce the severity of palmar-plantar erythrodysaesthesia syndrome (PPES; hand-foot syndrome).

Palliative radiotherapy was not permitted and concomitant use of medications that are strong inhibitors/inducers of CYP3A4 were to be avoided.</td></tr>
</table>

Source: CS Table 8 and Table 9

The CS provides a summary of the different data cut-off points used, combined with the outcome analyses and the source of the data (Table 5) and the key differences between the investigator and regulatory analyses of the trial (Table 6). As well as using different data cut-off points, the main differences between the two registration analyses appear to be the censoring rules and the use of one- or two-sided p-values (see Section 3.1.6 for more detail).

Table 5 CABOSUN data cut-off points and outcomes analysed

Date	Outcomes analysed	Source	Additional information
11 April 2016	PFS and (ORR)[a]	Choueiri 2016 et al[24]	Investigator assessment. Alliance censoring rules for progression (Missing or inadequate tumour assessments or use of systemic non-protocol anticancer therapy were not reasons for censoring) and 1-sided p-values. Event-driven analysis triggered when 123 events were observed.
15 September 2016	PFS and ORR	CSR and Choueiri et al 2017[30]	Additional analyses performed for regulatory purposes. Results in the CSR are based on assessment by an IRC and FDA-recommended censoring rules, with two-sided p-values. FDA-recommended censoring rules for PFS necessarily reduced the number of events available for analysis. To increase the number of events that would be included in the analyses, the data cut-off for radiographic endpoints in the CSR was extended to 15 September 2016 (database extract 13 January 2017 - the latest date for which OS data were available).
13 January 2017	OS (Exploratory analysis)	CSR	OS analyses were conducted with the most mature OS data available at the time.
1 July 2017	OS (Exploratory analysis)	Choueiri et al 2017[30]	Results from the updated OS analysis

Source: partly based on CS B.2.2 Table 6.
CSR, clinical study report; FDA, Food and Drug Administration; IRC, Independent Radiology Committee; ORR, objective response rate; OS, overall survival; PFS, progression-free survival.
[a] It is not clear why the objective response rate (ORR) next to progression free survival (PFS) in the first row of the CS table is bracketed.

As can be seen in Table 6, differences between investigator and regulatory analyses for CABOSUN resulted in different patient numbers for those with radiographic images and differences in the number of events recorded.

3.1.3.1 CABOSUN trial baseline characteristics

The CS states the demographic characteristics were well balanced between study groups, albeit as can be seen in Table 7, there are some exceptions. There are some differences in age range (cabozantinib 40-82 years; sunitinib 31-87 years), male sex (cabozantinib 84%; sunitinib 73%), prior nephrectomy (cabozantinib 72%; sunitinib 77%) and visceral metastases (cabozantinib 77%; sunitinib 72%). In response to a clarification request, the company confirmed that all RCC patients in the CABOSUN trial had metastatic disease (clarification question A2).

Table 6 Key differences between investigator and regulatory analyses for CABOSUN

	Original report (Choueiri 2016[24])	CSR and Choueiri 2017[30]	
Reader	**Investigator**	**Investigator**	**IRC**
No. of patients with radiographic images	157	157	156
No. of events	123	107	92
Cut-off date (PFS and ORR)	April 2016	September 2016	
Cut-off date (OS)	April 2016	January 2017 (CSR) / July 2017 (Choueiri 2017)	
Censoring rules (PFS)	Alliance	FDA guidance	
Censor for non-protocol systemic anticancer therapy	No	Yes	
Censor if event after ≥2 missing assessments	No	Yes	
Stratified analysis[a]	Yes	Yes	
P-value sided	1	2	

Source: CS B.2.2 Table 7
[a] Stratification factors: IMDC risk group (poor, intermediate) and bone metastases (yes, no).
CSR, clinical study report; FDA, Food and Drug Administration; IRC, Independent Radiology Committee; ORR, objective response rate; OS, overall survival; PFS, progression-free survival.

The ERG notes from CS Appendix D1.1 Table 11 that there were differences between the study groups in number of metastatic sites, with the percentage of patients with ≥3 sites 32% in the cabozantinib group, compared to 41% in the sunitinib group. The corresponding figures for 2 sites were 47% versus 26%, and corresponding figures for 1 site were 22% versus 33%. Thus, a greater proportion of patients in the cabozantinib arm had two or more metastatic sites (79%) than in the sunitinib arm (67%).

Clinical expert advice to the ERG suggests that these differences are not large enough to be of clinical importance. Expert clinical advice to the ERG also suggests the baseline characteristics are generally representative of patients seen in UK clinical practice apart from the proportion of patients with prior nephrectomy. This is higher than normally seen in clinical practice based on the experience of one of the experts.

3.1.3.2 Non-randomised trials

The CS for the clinical effectiveness of cabozantinib was limited to RCTs and no non-randomised studies were included in the submission.

Table 7 CABOSUN baseline patient characteristics

Characteristic, n (%)	Cabozantinib (n=79)	Sunitinib (n=78)
Age, median years (range)	63 (40-82)	64 (31-87)
Sex, male	66 (84)	57 (73)
Race		
White	70 (89)	75 (96.2)
Black	3 (4)	2 (2.6)
Asian	2 (3)	0
Other, unknown or not reported	5 (6)	1 (1)
ECOG PS		
0	36 (46)	36 (46)
1	33 (42)	32 (41)
2	10 (13)	10 (13)
IMDC risk group		
Intermediate	64 (81)	63 (81)
Poor	15 (19)	15 (19)
Bone metastases		
Yes	29 (37)	28 (36)
No	50 (63)	50 (64)
Prior nephrectomy		
Yes	57 (72)	60 (77)
No	22 (28)	18 (23)
Metastases[a]		
≥ 1 metastatic site	79 (100)	78 (100)
Visceral metastases	61 (77)	56 (72)

Source: CS Table 10

There is a small error in the CS table of baseline patient characteristics (CS Table 10), with the number of participants under race in the cabozantinib arm totalling to 80 rather than 79. It would appear that the number of Asian participants should have been one rather than two, as per the trial publication.[23]

[a], as reported by the investigator on the on-study case-report form. ECOG, Eastern Cooperative Oncology Group; IMDC, International Metastatic Renal Cell Carcinoma Database Consortium; PS, performance status.

3.1.4 Description and critique of the approach to validity assessment

The CS included a risk of bias assessment (CS Table 15 appendix D1.3) using the criteria suggested by NICE[31] for the CABOSUN and COMPARZ RCTs (details of the latter are reported in 3.1.7). Table 8 shows the company's and the ERG's quality assessment of the trial.

Table 8 Company and ERG assessment of trial quality - CABOSUN

NICE QA Criteria for RCT	CS response	ERG response
1. Was the method used to generate random allocations adequate?	Yes	Unclear risk of bias
Comments: Random stratified assignment [IMDC risk category (intermediate or poor) and presence of bone metastases (yes or no)] in a 1:1 allocation ratio using a dynamic allocation method. Dynamic allocation (DA) methods balance prognostic factors between treatment groups, which are a primarily deterministic, non-random algorithm.[32] However, DA is a family of methods, not just one, and the company does not specify which approach they used. The ERG requested details of the DA process employed (clarification question A4c), but the company did not provide any information beyond that already stated in the CS. It is therefore unclear why DA was needed given that there is already stratification, which prognostic variables were included in the DA algorithm and what part of the DA algorithm was random.		

2. Was the allocation adequately concealed?	Not clear	Unclear risk of bias
Comments: The method of allocation concealment is not reported in the trial publication, study protocol or CS.		
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Yes	Yes (low risk of bias)
Comments: The publication and the CS state that overall, the treatment groups were balanced with respect to baseline demographic and disease characteristics. However, there were minor differences between the treatment arms, with the cabozantinib arm containing 11% more male patients, a slightly different ethnic mix (7.6% fewer white patients), 5% fewer patients who had had a prior nephrectomy, and 5% more patients with visceral metastases than the sunitinib arm. A greater proportion of patients in the cabozantinib arm had two or more metastatic sites (79%) than in the sunitinib arm (67%). Clinical expert advice to the ERG suggests that these minor differences would be unlikely to have clinical implications.		
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	No	No (high risk of bias)
Comments: Open label trial. The CS states that a central imaging review of investigator assessments was not performed. However, a blinded central review by an IRC was undertaken retrospectively to minimise bias for the PFS and response outcomes in the company's updated analysis.		
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No	Unclear risk of bias
Comments: The most frequent primary reasons for study treatment discontinuation were disease progression (cabozantinib 56%, sunitinib 53%) and AEs (20% and 21%, respectively) (clarification question A15). The company states that in general, the numbers of dropouts were considered balanced, and the ERG agrees that this is the case for withdrawal due to progression and AEs. However, there were differences between the study arms in the number of patients who did not receive the study drug (cabozantinib n=1, sunitinib n=6) and in the number of patients who withdrew consent (cabozantinib n=3, sunitinib n=7 according to Figure 52 in CS Appendix D1.2; but n=1 and n=9 respectively according to CS section B.2.13). The company states that the frequency of withdrawal by subject during treatment is considered as low (clarification question A15). We note that these withdrawals amount to 3.8% of the cabozantinib trial arm and 9.0% of the sunitinib trial arm. It is unclear whether this difference would have introduced bias, since the reasons for patients' withdrawal of consent are not reported. It should also be noted that there was an imbalance between the cabozantinib and sunitinib arms in the proportions of patients who had ≥2 missed "adequate tumour assessments" before a PFS event, and in the proportions who had no post-baseline "adequate tumour assessments". In response to a clarification request (question A7), the company states that the reasons for these differences are not available.		
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No (low risk of bias)
Comments: There are no deviations from the trial protocol with regard to outcomes.		
7. Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes (low risk of bias) Yes Yes
Comments: States that an ITT approach (defined as all patients who were randomised) was used for all but safety data (the safety analysis population was defined as patients who received ≥1 dose of study drug). In response to a clarification request on missing data, the company states that in the retrospective IRC assessment of PFS and ORR, no values were imputed for patients for whom a complete set of baseline and post-baseline radiographic images were not available and FDA censoring rules were applied (clarification question A6). The application of FDA-recommended censoring rules for PFS necessarily reduced the number of events available for analysis (CS Section B.2.2). The CS states that in the retrospective IRC assessment of PFS and ORR, no values were imputed for patients for whom a complete set of baseline and post-baseline radiographic images were not available (CS Table 11). Therefore 156 patients with radiographic images and 92 events were included in the		

retrospective analysis compared to 157 patients and 123 events in the original analysis (CS Table 7).

The ERG's quality assessment mostly agrees with that of the company. The ERG disagrees with the company that there is no risk of bias for random sequence generation and for allocation concealment. In the ERG's view the risk is unclear as adequate information has not been provided on procedures. Both the company and the ERG agree that the trial is at a high risk of bias due to being open-label. However, a blinded retrospective review by an independent radiology committee (IRC) was undertaken to minimise detection bias for the PFS and response outcomes in the company's updated analysis. Overall, the ERG is of the opinion that the CABOSUN trial appears to have been well conducted though with some limitations as outlined above.

3.1.5 Description and critique of company's outcome selection

The outcomes in the CS match those listed in the NICE scope and the decision problem. These are:

- PFS - defined as the interval between randomisation and first documentation of disease progression, or death from any cause. This outcome was originally investigator-assessed. For the regularity submission, a blinded, retrospective central review of the radiographic images was carried out by an IRC to determine progress and response. The CS presents IRC-assessed results for this outcome. Progression was assessed according to RECIST 1.1 at screening and every two treatment cycles (i.e. every 12 weeks).
- OS - defined as time from randomisation to death from any cause.
- ORR - defined as the proportion of patients at the time of data cut-off with a best overall response of CR (complete response) or PR (partial response), confirmed by a subsequent visit ≥ 28 days later (assessment as for PFS).
- Adverse events - graded by Common Terminology Criteria for Adverse events (CTCAE) version 4. Safety was assessed on a schedule based on the date of the first dose, days 15 and 29 of Cycle 1 and 2, and day 1 of each subsequent cycle.

The above outcomes are valid and appropriate endpoints used in cancer trials. Of these, only ORR is not used in the economic model of the CS.

In addition to the listed outcomes, the company states 'Duration of response' under 'all other reported outcomes' (CS Table 8). No definition for this outcome is provided.

HRQoL data were not collected in the CABOSUN trial and hence not reported for the clinical effectiveness section of the CS. Phase II clinical trials generally do not assess outcomes such as HRQoL. HRQoL in cancer trials it is an important outcome that should be included, as it generally reflects a patient's day-to-day functioning.[33] For the economic model, the company used other published sources of HRQoL data, as discussed in section 4.3.5 of this report.

3.1.6 Description and critique of the company's approach to trial statistics

The CS reports results for all of the outcomes specified in the NICE scope, apart from HRQoL which had not been assessed in the CABOSUN trial (CS Table 1).

The statistical analysis approaches employed in the CABOSUN trial are summarised in CS Table 11. The CSR states that the statistical analysis plan for CABOSUN is available in an Appendix of the CSR; this was not available to the ERG and was requested by the ERG from the company (clarification question A20).

3.1.6.1 Statistical analysis approaches

Two different analysis approaches were employed in the CABOSUN trial:

- the original analysis, as reported in the CSR and the trial publication;[24]
- an updated analysis that was conducted by the company to meet regulatory requirements (CS Table 7).

The CS states that the company's submission to NICE is based on the updated analysis and therefore results as reported in the CS differ in some respects to those reported in the trial publication (CS section B.2.2).[24] Results of the updated analysis are also reported in the CSR and in a conference presentation.[30]

Standard statistical methods were used to compare time-to-event outcomes between cabozantinib and sunitinib (CS section B.2.4). Kaplan-Meier (K-M) curves are presented in CS Figure 5 for PFS and in CS Figures 6 and 7 for OS. The hazard ratios were estimated based on Cox regression with a 2-sided log-rank test stratified by IMDC risk group (poor, intermediate) and bone metastases (yes, no) (for a definition of the IMDC risk factors see section 2.1.4). The CS clearly reports the number of patients at risk at each time point; the number of patients censored for in each trial arm, with reasons (CS Table 12 for PFS; CS Table 13 for OS); the median PFS and OS with 95% confidence interval for each trial arm; the hazard ratio (HR) with 95% confidence interval; and the p-value from the log-rank test (CS Figure 5 for PFS; CS Figures 6 and 7 for OS).

3.1.6.2 PFS (primary outcome)

The original analysis approach for PFS, as reported in the trial publication,[24] employed unblinded radiological assessments made by the trial investigators, censoring according to Alliance rules (missing or inadequate tumour assessments or use of systemic non-protocol anticancer therapy were not censored), a one-sided hypothesis test, and a data cut-off of April 2016. The company's updated analysis, as presented in the CS, required radiological assessments to be made retrospectively by a blinded IRC, censoring according to FDA rules (missing or inadequate tumour assessments or use of systemic non-protocol anticancer therapy were censored), a two-sided hypothesis test, and was based on a data cut-off of September 2016 (CS Table 7). Median follow-up for PFS in the updated analysis was 25.0 months.

The data cut-off for progression in the original analysis was event-driven, with analyses being triggered when 123 events were observed. For the updated analysis, the CS states that to increase the number of events that would be included in the analyses, the data cut-off for radiographic endpoints was extended to 15th September 2016. We note that the number of events achieved at this later cut-off (CS Table 7) was less than the 123 specified in the power calculation (see 'Sample size and power calculation' below).

3.1.6.3 OS (secondary outcome)

The original analysis of OS, as reported in the trial publication,[24] was based on a data cut-off of April 2016. The updated analysis, reported in the CSR and CS, employed a data cut-off of 13 January 2017, with a median follow-up of 28.9 months. The CS also reports an analysis of OS at the latest available data cut-off, 1 July 2017 (as reported in a conference presentation[30]) (CS Table 7). Median follow-up was not reported for this analysis.

The OS data at all the analysis time points were immature. The CS cautions that there was a notable degree of censoring around the median estimates, and confidence intervals around the hazard ratios were wide due to the relatively low number of deaths (CS section B.2.6).

The CSR states that "the study did not have a pre-specified hypothesis for the treatment effect on OS, so inference tests should be interpreted accordingly" (CSR section 11.5).

3.1.6.4 ORR (secondary outcome)

The initial and updated analysis approaches for ORR were the same as those employed for PFS (CS Table 7).

Standard statistical methods were used to compare the ORR between cabozantinib and sunitinib (CS section B.2.4). The difference in percentage ORR between groups was tested with a 2-sided Cochran-Mantel Haenszel (CMH) test with the same stratification factors as the PFS analysis. The CS clearly reports the percentage ORR with 95% confidence interval for each trial arm; the ORR treatment difference with 95% confidence interval; and the p-value for the difference from the CMH test (CS Tables 14 and 15).

In addition to the ORR, the CS reports descriptively (i.e. without statistical analysis): the numbers and percentages of patients in each trial arm with: a complete response; a confirmed partial response; stable disease; progressive disease; unevaluable or missing data; the percentage with any reduction in the target lesion; and the disease control rate (CS Tables 14 and 15). According to footnote d in CS Table 15, the CS reports the percentage with progressive disease as "progressive disease as best overall response". The company clarified that this refers to the proportion of patients whose best overall response to treatment with regard to tumour response was classified as 'progressive disease' (clarification question C1).

3.1.6.5 Analysis populations

The CS states that all efficacy analyses were carried out in the intent-to-treat (ITT) population, defined as all patients who were randomised. The safety analysis population was defined as all patients who received any treatment with cabozantinib or sunitinib. Patients were analysed according to actual treatment received (CS Table 11).

3.1.6.6 Sample size and power calculation

The null hypothesis in the initial analysis of PFS was that the HR for progression of the cabozantinib and sunitinib arms would be 1.0. The alternative 1-sided hypothesis was that the HR would be 0.67, favouring cabozantinib over sunitinib.

The CS reports that a sample size of 123 events (progressions or deaths) would provide the log-rank test with 85% power to detect a HR of 0.67 for PFS, assuming a 1-sided type I error rate of 0.12, equivalent to an increase in median PFS from 8 months in the sunitinib arm to 12 months in the cabozantinib arm (CS Table 11). Assumptions required to achieve the target of 123 events are stated in CS Table 11 (including 5.8% accrual rate over 24 months, minimum PFS follow-up 20 months, and exponential distribution of PFS).

We note that the updated analysis of PFS as reported in the CS would have been under-powered statistically compared to the initial analysis specified in the sample size calculation, since a 2-sided test has less statistical power than a 1-sided test, and 92 events occurred in

the updated analysis due to different censoring rules, which is fewer than the planned target of 123 events (CS Table 7).

3.1.6.7 Treatment of missing data

As noted above, the CS states that censoring rules for the updated analyses of PFS were applied in accordance with FDA guidance (CS Table 11); the FDA rules can be inferred from CS Table 12 and are stated explicitly by the company in their response to clarification question A6. In the retrospective IRC assessment of PFS and ORR, no values were imputed for patients for whom a complete set of baseline and post-baseline radiographic images were not available (CS Table 11).

The reasons for censoring PFS data in the retrospective IRC analyses, based on FDA rules, were: ≥2 missed analyses prior to an adequate tumour assessment (ATA); no baseline and post-baseline ATA; no event by the last ATA; no post-baseline ATA, and receipt of systemic anticancer therapy (CS Table 12). As noted above, missing or inadequate tumour assessments or use of systemic non-protocol anticancer therapy were not reasons for censoring in the initial investigator analysis approach using the Alliance censoring rules (CS Table 7).

According to CS Table 12, there were imbalances between the cabozantinib and sunitinib arms in the proportions of patients who had ≥2 missed adequate tumour assessments before a PFS event (6% versus 0%) and in the proportions who had no post-baseline adequate tumour assessments (1% versus 8%). The company explained in a clarification response that information on the reasons for these differences is not available (clarification question A7).

The CS states that there was an imbalance in the number of patients with missing data (CS section B.2.13). One patient in the cabozantinib arm and six in the sunitinib arm withdrew prior to receiving study treatment, but the reasons for these withdrawals were not known. There was also a higher incidence of missing or unevaluable data in the sunitinib arm, with six patients in the cabozantinib arm and 18 in the sunitinib arm not evaluable because they had no adequate post-baseline imaging assessments. The reasons were: cabozantinib: adverse event (n=5), withdrew consent (n=1); sunitinib: adverse event (n=6), death (n=2), disease progression (n=1), withdrew consent (n=9). We note that the numbers who withdrew consent are slightly different in CS Appendix D.1.2 Figure 52, which gives 3 and 7 in the cabozantinib and sunitinib arms respectively. The CS states that “because of the nature of these clinical events, none of these patients was likely to have experienced a response or

prolonged PFS". The CS further states that "based on their baseline characteristics (unavailable to the ERG to verify), the sunitinib patients without post-baseline imaging would not be expected to have a better prognosis than sunitinib patients who had a response recorded, and therefore it is unlikely that the radiographic endpoints were biased against sunitinib by these missing data" (CS section B.2.13). Clinical experts advising the ERG suggested that whilst this assumption may be reasonable, it is difficult to be sure (given the lack of data on the characteristics of patients with and without post-baseline imaging). Experts also commented that the 9 patients who withdrew consent in the sunitinib arm is a relatively high proportion (i.e. 11.5% of patients in the sunitinib arm) and, speculatively, might reflect their dissatisfaction with assignment to the comparator rather than to the experimental treatment. However, we note that an imbalance in the number of patients who withdrew consent was not seen in the open-label COMPARZ trial, where 6.6% and 6.7% of patients in the pazopanib and sunitinib arms withdrew consent.[34] We also note an unexplained inconsistency in the number of patients who withdrew consent in the CABOSUN trial, as reported in the CS, which differs between CS section B.2.13 (1 and 9 withdrew from each trial arm) and CS Appendix D.1.2 Figure 52 (3 and 7 withdrew).

There appears to be inconsistency in the CS regarding the number of inadequate radiographic images or tumour assessments. CS section B.2.6 states that 13 patients did not have complete data for radiographic images or tumour assessments but these do not appear to have been accounted for among the 24 patients mentioned in CS section B.2.13 (as referred to above), who did not have adequate post-baseline imaging assessments. Further, CS Table 14 suggests that the number not evaluable was 10. The company clarified that these differences are due to the timing of the assessments for ORR responses being the entire period prior to progression, while for PFS only the response at time of progression was considered. The difference in the patient numbers seen thus reflects the fact that ORR and PFS were mostly evaluated at different numbers of points (clarification question A8).

Sensitivity analyses were conducted to explore the effect of potentially informative censoring in PFS analyses based on IRC assessments; these are reported in the CSR but not in the CS. Four sensitivity analyses examined the impact of (a) discontinuation of study treatment for reasons other than radiographic progression with no non-protocol anticancer therapy (NPACT) or (b) receipt of NPACT prior to progression. The four analyses were (i) censored subjects meeting criterion (a) were classified as events in both treatment arms; (ii) censored subjects meeting criteria (a) or (b) were classified as events in both treatment arms; (iii) censored subjects in the cabozantinib arm meeting criterion (a) were classified as events in the cabozantinib arm but remained censored in the sunitinib arm; and (iv) censored subjects

in the cabozantinib arm meeting criteria (a) or (b) were classified as events in the cabozantinib arm but remained censored in the sunitinib arm (this was the most conservative analysis) (CSR Table 17).

3.1.6.8 Subgroup analyses

Subgroup analyses of PFS per IRC assessment, with censoring according to FDA rules, are mentioned briefly in CS section B.2.7 and are presented in CS Appendix E. Subgroup analyses were pre-planned, except for age, race and sex which were exploratory analyses (CS Table 9). A total of 16 HRs for cabozantinib versus sunitinib are presented in a forest plot for the following subgroups:

- The analysis stratification factors: IMDC risk category (referred to as "Heng risk factors" in CS Appendix Figure 53) (intermediate, poor); and bone metastases (yes, no);
- MET status (positive, negative, missing);
- Age, years (<65, ≥65);
- Sex (male, female);
- Race group (white, other);
- Baseline ECOG performance status (0, 1, 2).
- Bone metastases (yes, no)

For each subgroup, CS Appendix Figure 53 presents the number of events and the median PFS in in each trial arm, and the HR with 95% confidence interval. The CS, CSR and Statistical Analysis Plan do not specify whether an adjustment was made to the type I error rate to account for multiple subgroup testing. The company confirmed in a clarification response that no adjustment was made (clarification question A9). The CSR states that, for completeness, HRs (and 95% CIs) were generated regardless of the size of the subgroup (CSR section 11.4.3.9). Subgroup sizes ranged from 8 to 70 subjects in the cabozantinib arm and from 3 to 75 subjects in the sunitinib arm (CS Appendix E Figure 53).

The CSR and Statistical Analysis Plan report that further subgroup analyses of OS, ORR and PFS were conducted per investigator radiology assessment and following both Alliance and FDA censoring rules, although these analyses are not included in the CS. The company provided subgroup analysis results for OS in response to a request by the ERG (clarification question A9).

3.1.6.9 Summary of company's approach to trial statistics

Overall, the ERG agrees with the company's approach to statistical analysis, which employed standard methods. We also agree with the company's caution that the OS data at all time points are immature and should be interpreted with caution. The key limitation in the company's approach noted by the ERG is that there were unexplained imbalances between the trial arms in missing data on tumour assessments and in patient discontinuations due to withdrawal of consent, and it is unclear whether these might have introduced bias. We also note that the updated analysis of PFS is under-powered relative to the power specified in the sample size calculation. Subgroup analyses included some subgroups with small sample sizes and no adjustment was made to control the type I error rate when analysing multiple subgroups.

3.1.7 Description and critique of the company's approach to the evidence synthesis

The CS presents a narrative review of clinical effectiveness, with study characteristics and results presented in text, tables and figures. As only one RCT of cabozantinib was included in the systematic review a meta-analysis of cabozantinib trials was not possible. However, to facilitate comparison with pazopanib an indirect treatment comparison (ITC) was performed, for the outcomes of PFS and OS. The following sections describe and critique the ITC, and a tabulated critical appraisal can be found in Appendix 9.1.

3.1.7.1 ITC evidence networks

The CS reports that a total of 19 trials (n=105 records) were identified for inclusion in the ITC (CS Section B.2.9), based on criteria that included treatments for RCC within the NICE scope (cabozantinib, sunitinib, pazopanib) and treatments outside the scope (interferon alfa, interleukin-2, sorafenib, bevacizumab and interferon alfa, temsirolimus, tivozanib, placebo). The ERG notes from CS table 18 that a total of 13 RCTs (reported in 19 publications) were included in this network, which is a discrepancy with the reported 19 RCTs mentioned in the CS.

CS Figure 9 illustrates the evidence network constructed from the 13 RCTs (reproduced below in Figure 4 – it is not stated whether this network is specific to OS or PFS outcomes, or both). In this network cabozantinib is connected via sunitinib (from the CABOSUN trial), which in turn is connected to sorafenib, pazopanib, interferon alfa, and bevacizumab and interferon alfa. These treatments in turn connect to tivozanib, placebo, and temsirolimus. For some comparisons the network contains both direct and indirect evidence ("closed loops"), and for other comparisons only direct evidence is included. The CS refers to this as a potential evidence network constructed to identify additional connections between

cabozantinib and pazopanib. Hereafter the ERG refers to this as the "wider network" of 13 RCTs (i.e. containing both in-scope and out-of-scope treatments).

The CS subsequently restricted inclusion to the ITC only to studies which included the comparators relevant to the scope of the appraisal (pazopanib and sunitinib). Studies which did not include these comparators were excluded unless they provided an intermediate link. The "restricted network" included two studies: CABOSUN (comparing cabozantinib with sunitinib) and COMPARZ (comparing sunitinib with pazopanib).[34 35] The restricted evidence network therefore includes three treatments connected via a common comparator, sunitinib (CS Figure 11 reproduced below in Figure 5).

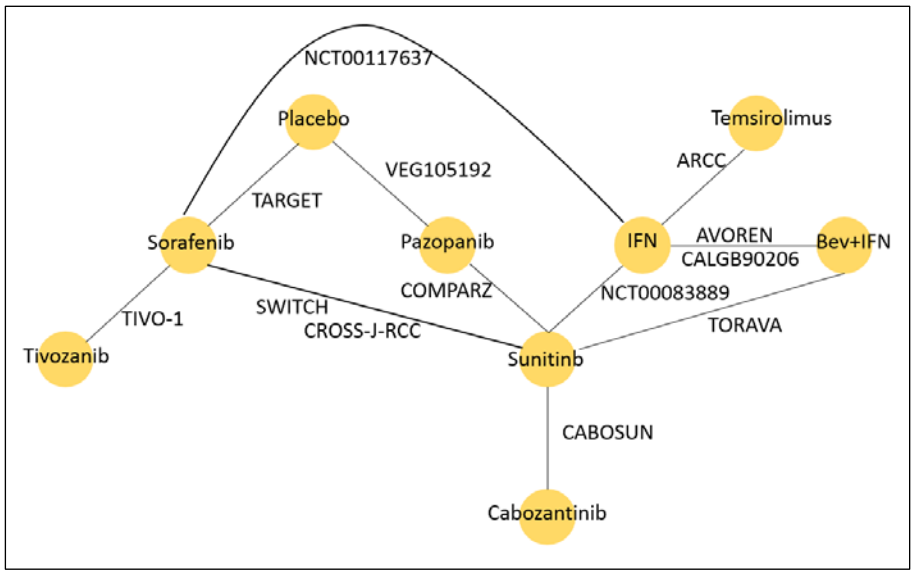


Figure 4 Wider evidence network of 13 trials (reproduced from CS Figure 9)

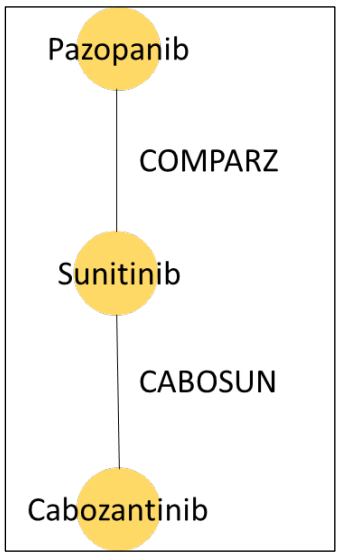


Figure 5 Restricted evidence network (reproduced from CS Figure 11)

The CS does not provide a heterogeneity assessment (statistical or clinical) of the trials in the wider network of 13 RCTs, and does not report results of any ITC based on this network. The company was requested to provide ITC results using this wider network (clarification question A26) to permit comparison of the results of the wider network with the restricted network (i.e. to check whether the results for the comparison between cabozantinib, sunitinib and pazopanib were different when a wider network containing other treatment comparisons was used). The company provided these results as survival curves, HR plots and tabulated HRs, for OS and PFS, for two analysis approaches which they had used to conduct the ITCs - Ouwens et al[36] parametric survival models and fractional polynomial models (see section 3.1.7.3 below for an explanation of these models), based on both random effects and fixed effects.

ITC feasibility assessment

The CS reports conducting a feasibility assessment for the ITC (CS section B.2.9). This assessment had two stated components: to assess whether adequate outcome data were available; and to assess whether there were differences in study and patient characteristics within and between treatment comparisons that might influence treatment effects (i.e. clinical heterogeneity). This feasibility assessment appears to have been applied only to the two trials included in the restricted ITC network (i.e. not to the wider network of 13 RCTs described above).

CS Tables 20 and 21 report the data availability assessments for PFS and OS, respectively. Hazard ratios for the ITT population (adjusted/stratified and unadjusted/unstratified) and RCC risk subgroups (intermediate risk and poor risk) are tabulated for both trials. The CS states that PFS data would be acceptable if measured either by IRC or by study investigators, with the IRC assessment considered by the company to be less likely to be biased and prioritised where possible.

Clinical heterogeneity

The CS states that the differences in distribution of RCC risk category is the variable that most affects survival (CS Appendix D). CS Table 22 compares risk category and ECOG performance status between the two included trials. The CABOSUN trial classified risk status according to the IMDC criteria (for definition of these see section 2.1.4), whilst the CS states that the COMPARZ trial used the original MSKCC criteria. However, Table 11 in Appendix D1.1 reports both IMDC and MSKCC risk classifications for the COMPARZ trial.

The ERG notes that the distribution of patients across risk categories for these two instruments in this trial are broadly similar. Expert clinical advice to the ERG is that MSKCC and IMDC are similar, thus differences between the trials in how patients were classified would be unlikely.

The CABOSUN trial included only patients at intermediate or poor RCC risk, whilst the COMPARZ study included patients with favourable, intermediate and poor risk classifications. The distribution of patients between risk classifications is therefore different between the two trials. Approximately 80% of patients in the CABOSUN trial were at intermediate risk, compared to approximately 54% to 56% in COMPARZ, and approximately 19% of patients were classified as poor risk in CABOSUN compared to 17 to 19% in COMPARZ (all figures based on the IMDC risk classification). The percentage of patients with favourable risk in COMPARZ was 25%, with no favourable risk patients in CABOSUN for the reason stated above. The patient RCC risk profile in COMPARZ is therefore more favourable than in CABOSUN. The CS does not comment on the impact of this difference, but the ERG considers this would likely under-estimate the relative effectiveness of cabozantinib compared to pazopanib in the ITC since patients in the COMPARZ trial overall have a lower RCC risk and accordingly could be expected to respond more favourably to treatment.

Cancer performance status was reported by ECOG classification in CABOSUN and the Karnofsky index in COMPARZ. In CABOSUN around 46% of patients were classified as EGOG 0 (which indicates the patient is fully active, and able to carry on all pre-disease performance without restriction), and around 41% were classified as ECOG 1 (which indicates mild restriction in ability to carry out physical activity and work). In COMPARZ around 75% of patients had a Karnofsky score of 90 to 100%, indicating normal activity, no/minor signs of disease (NB. The data for Karnofsky performance status 70 to 80 and 80 to 100 are the wrong way round in CS Table 11). An ECOG performance status of 0 is considered comparable to Karnofsky score of 90% to 100%, and an ECOG performance status 1 is comparable to a Karnofsky score 70% to 80%.[37] Thus, the two trials are broadly comparable in terms of cancer performance status, though it appears that a greater proportion of COMPARZ patients were classified as having the highest performance status. Expert clinical advice to the ERG is that this is likely to be due to some of the patients in COMPARZ having favourable risk status (ECOG performance status is one of the constituent variables in the risk status assessment).

There were slight differences between trials in the number of metastatic sites detected (≥3 sites: 32% to 41% by treatment arm in CABOSUN; 42% to 44% by treatment arm in COMPARZ). (CS Appendix Table 11). Just over a third of patients in CABOSUN had bone metastases at baseline (36% to 37% by trial arm) compared to 15% to 20% (by trial arm) of patients in COMPARZ. The CS states that patients with bone metastases have a poor prognosis and experience poorer outcomes with currently available treatments compared with patients without bone metastases. A higher percentage of patients in COMPARZ received prior nephrectomy (82% to 84% by trial arm) compared to patients in CABOSUN (72.2% to 76.9% by trial arm). Expert clinical advice to the ERG suggests this may be explained by the fact that patients with more favourable RCC risk are more likely to receive nephrectomy (COMPARZ included some patients with favourable RCC risk). Also, fewer nephrectomies tend to be performed now in practice than in the past (COMPARZ is an older trial than CABOSUN). Expert clinical advice also notes that prior nephrectomy is associated with a better treatment outcome, thus raising the potential risk of bias in the ITC results.

There were differences in ethnicity between the two trials: 92% of patients were classified as white in CABOSUN, compared to 64% white in COMPARZ (34% were described as being Asian). All patients in CABOSUN were from the USA, whereas patients in COMPARZ were from 14 countries located in North America, Europe, Australia, and Asia. The trials were comparable in terms of age (median age 61 to 64 across the trials) and reasonably similar in gender profile (male: 73.1% to 83.5% in CABOSUN; 71% to 75% in COMPARZ). The inclusion criteria of both trials required patients to have locally advanced or metastatic clear cell RCC. All patients in the CABOSUN trial had metastatic disease, whilst 98% had stage IV disease in COMPARZ.

In terms of design characteristics, the CABOSUN trial was a phase II RCT (n=157 patients, of whom 79 were randomised to cabozantinib and 78 were randomised to sunitinib), whilst COMPARZ was a larger phase IIIb non-inferiority RCT (n=1110 patients randomised, of whom 557 were randomised to pazopanib and 553 were randomised to sunitinib). The ITC is therefore unbalanced in terms of the proportions of patients randomised to the three respective treatments. The primary outcome measure in both trials was PFS.

In both trials the study treatments were administered continuously until progression of disease, the occurrence of unacceptable toxic effects, or withdrawal of consent. The dose regimen of sunitinib was identical in both trials (orally once per day at a dose of 50 mg for four weeks, followed by a two-week break).

Patient crossover was not permitted in the CABOSUN trial, and the CS states that the occurrence of crossover was not reported in the COMPARZ trial (CS Appendix D1.1 Table 10). The ERG has checked the available reports of the COMPARZ trial and can find no mention of crossover.[34 35 38 39] As reported earlier, 57 to 58% of patients in the CABOSUN trial received subsequent anticancer drug treatments following discontinuation of study treatment. In COMPARZ 55% to 56% of patients received subsequent anticancer therapy, including sunitinib in pazopanib-treated patients, and vice versa. The occurrence of subsequent anticancer treatment will affect estimates of OS in both trials.

In summary, there some similarities but also a number of differences between the two RCTs in the ITC, with the most important difference being in RCC risk status. Overall, patients in the CABOSUN trial had a poorer RCC risk status and cancer performance status than patients in the COMPARZ trial. The CS does not comment on the likely implications of this on the ITC results. The ERG considers the effect of this on the ITC results to be uncertain.

3.1.7.2 Critical appraisal of trials included in the ITC

CS appendix D provides the company's critical appraisal of the two trials included in the ITC (Figure 41 and 42 and Table 15). A brief commentary is provided in which it is stated that the trials met assessment criteria for method of randomisation, balanced trial arms at baseline, no selective reporting and use of ITT analysis. However, it is stated there was potential risk of bias due to lack of patient blinding to treatment allocation (both trials were open label), and lack of information on allocation concealment. As discussed earlier in this report (section 3.1.4), the ERG mostly agrees with the company's critical appraisal of the CABOSUN trial. The ERG also conducted an independent critical appraisal of the COMPARZ trial to compare with that of the company (see Appendix 9.2). The ERG notes that a blinded central review by an IRC was undertaken in both trials for the PFS and response outcomes (retrospectively in CABOSUN), thus the potential for detection bias is reduced for those outcomes, though performance bias is still possible.

In summary, the ERG considers the methodological quality of the two trials to be adequate overall and the overall risk of bias to be low, with the exception of bias related to lack of blinding, and bias relating to sequence generation and allocation concealment procedures which were not clearly reported. The other limitations of the CABOSUN trial need to be acknowledged, namely, the fact that it is a relatively small phase II trial with immature OS data.

3.1.7.3 Statistical ITC methods used

Three different statistical methods were used to conduct the ITC:

(1) Indirect comparison of parametric survival curves using methodology developed by Ouwens et al (2010).[36]
(2) Parametric models with fractional polynomial distributions using methodology developed by Jansen (2011).[40]
(3) A "network meta-analysis: supplementary method" comparing hazard ratios using a fixed effects model, for intermediate risk and poor risk subgroups and the ITT population.

Methods 1 and 2 were used to inform the economic model and are included in the CS due to the company's observation that the assumption of proportional hazards was violated in the CABOSUN trial for OS and PFS, and for PFS in COMPARZ (Appendix D1.1 Table 12). The ERG concurs that proportional hazards do not hold for OS in CABOSUN as the survival curves in CS Figure 6 clearly cross at around month 14. However, the PFS survival curves (CS Figure 5) appear parallel after around month three. In the COMPARZ trial the ERG concurs that proportional hazards do not appear to hold for PFS based on visual inspection of the survival curves.[34 35] However, the ERG notes that the OS survival curves in this trial appear to cross at around month 24.[35] Because of these differences in opinion the company were asked to clarify their conclusions on the proportional hazards assumptions (clarification question A3).

The company responded by supplying scaled Schoenfeld plots and log-cumulative hazard plots for OS and PFS in both trials. Plots of Schoenfeld residuals against time are a standard approach to test for the (non-)proportionality of hazards; violation of the proportional hazards assumption is indicated if the plot of Schoenfeld residuals against time shows a non-random pattern. The company state that the Schoenfeld plots show an "increasing trend followed by a decreasing trend" and that the log-cumulative plots show "roughly parallel curves". The ERG considers that proportional hazards hold for PFS but not OS in the CABOSUN trial based on inspection of the log-cumulative hazard plots. For the COMPARZ trial the reverse is apparent: proportional hazards do not appear to hold for PFS but they do for OS. Given the observation of non-proportionality of hazards in at least one of the outcomes in both trials the ERG considers use of ITC methods that accommodate time-varying HRs to be appropriate.

Method 3 is presented as an additional analysis to explore comparative treatment effects in RCC risk groups. It does not assume proportional hazards and does not inform the economic model. We provide a brief description and critique of this analysis in Appendix 9.3.

The following sub-sections describe and critique, in turn, methods 1 and 2.

3.1.7.4 ITC: comparison of parametric survival curves

The CS reports use of a Bayesian statistical method described by Ouwens et al (2010) as a method for conducing an ITC.[36] This method was developed as an alternative to methods of assessing treatment effects which assume proportional hazards. The application of a constant HR implies the assumption that the treatment only has an effect on the scale parameter of a distribution. The method devised by Ouwens et al[36] uses parametric survival distributions to extrapolate outcomes which can be described by two parameters (shape and scale). The time-varying HR is expressed as a difference in scale and a difference in shape of the hazard functions of compared interventions. Ouwens et al[36] consider that encompassing treatment effects on both shape and scale is a more flexible approach to model relative survival. The method can be applied to pairwise meta-analysis of survival curves as well as multiple indirect comparisons of interventions. The similarity and consistency assumptions need to be fulfilled as they would do in other types of indirect comparison (see below).

The method can be used with both individual patient data and aggregated data from Kaplan-Meier curves. Scanned survival curves can be divided into multiple consecutive intervals over the trial follow-up period, and extracted survival proportions can be used to calculate the incident number of deaths for each interval and patients at risk at the beginning of the interval.[36]

Five parametric models were used by the company in the application of this method, four of which assumed two-parameter distributions (Weibull, log-logistic, log-normal, Gompertz), and one which used a one-parameter (exponential) distribution. The CS states that the exponential model was chosen because it made the same assumption as the previous method of hazard proportionality and allowed comparison. Model fit was assessed using the deviance information criteria (DIC) (CS Table 23).

Bayesian models were fitted using sunitinib as the reference treatment, and estimated treatments in terms of their effect on the reference parameters. The CS states that effect transitivity is an underlying model assumption. The transitivity assumption (also known as

the consistency assumption) requires covariates that act as relative treatment effect modifiers to be similar across trials. As discussed above, this assumption may not hold given the differences between the two trials in factors such as baseline RCC risk status and proportions of patients with bone metastases.

The parameter estimates for differences between treatments in scale and shape can be reported (accompanied by credible intervals), and expressed visually as HR and hazard rate plots showing treatment curves over the follow-up period.[36] The CS does not present hazard ratio or hazard rate plots, but does present fitted survival curves for the outcomes of OS and PFS for all three treatments, based on fixed effects and random effects models, for each of the five parametric survival distributions (Figure 1 to Figure 20, Appendix D1.1).

In summary, the ERG considers the Ouwens et al[36] method appropriate for implementing the ITC given the violation of the proportional hazards assumption for OS in the CABOSUN trial (notwithstanding the aforementioned caveats about clinical heterogeneity between trials). The ERG notes that this method has been used in two previous NICE appraisals, of breast cancer treatment (TA239 and TA503).[41 42]

3.1.7.5 Fractional polynomial model

The CS cites a publication by Jansen[40] as the basis of their use of fractional polynomial methodology. Jansen describes this method as an alternative to NMA of survival data in which the treatment effect is represented by a constant HR. A multi-dimensional treatment effect approach is used in which hazard functions of interventions compared in an RCT are modelled, and the difference between the parameters of these fractional polynomials within a trial are synthesized (and indirectly compared) across studies. The fractional polynomial analysis generates results which reflect the time course of the log-hazard function and as such can be expressed as log-hazard function curves and their parameters (intercept and slope). Credible interval curves can be plotted alongside the log-hazard function curves. The ERG notes that fractional polynomial-based NMAs have also been included in other NICE STAs, including appraisals of renal cell carcinoma treatments (TA463 and TA512).[14 22]

Two orders of FP model were considered for inclusion: first-order, and second order. The power level for each order can be chosen from the following set -2. -1, -0.5, 0, 0.5, 1, 2, 3. A first order model with a P= 0 would be equivalent to a Weibull model, and a first order model with P=1 would correspond to a Gompertz model. For the first order model the following powers were considered in the CS: P=-1, P=-0.5, P=0, P=0.5 and P=1. For the second order

model the following powers were considered: P1=-0.5, P2=0; P1=-1, P2=0; P1=-1, P2=-1; P1=-1, P2=0.5; P1=-1, P2=1 (see CS Table 24).

The ERG notes that only a relatively narrow range of powers (P1 and P2 in the range -1 to +1) were considered in the company's analysis. Given that none of the modelled OS curves in the CS appeared to reflect the shape of the CABOSUN KM OS curves, the company was asked if they had considered a wider range of powers (thus reflecting other functional forms) (clarification question A24). The company responded with a number of justifications for their chosen range. They stated that the joint estimation of parameters "is very delicate for every (P1, P2) model and the lack of stability of the estimation algorithms typically causes very long run times" thus they had to be strategic in their choice of which powers to test. They also cite their previous submission to NICE on cabozantinib for second line treatment of RCC and the fact that the best fitting fractional polynomial in that submission was within the same range of powers. They also state that their guiding principle was that smaller values of P1 and P2 should be preferred, implying that using higher power values would lead to over-fitting which would give curves uncharacteristic of typical PFS or OS curves. Overall, the ERG considers that the justification given by the company for the range of powers tested is reasonable.

3.1.7.6 Choice of fractional polynomial model

To select the most appropriate fractional polynomial model from the first and second order models considered, the company used the DIC to compare goodness-of-fit. The DIC is commonly used to compare the fit of Bayesian statistical models. The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset which has the same structure as that currently observed.[43] The best fitting fractional polynomial model chosen for OS and PFS was the second order model P1=-1 and P2=-1 (CS Table 24), and this was used to inform the economic model (CS Table 59). The CS does not state whether any other considerations were taken into account in the choice of model, such as clinical plausibility with respect to the OS and PFS estimates generated. They comment that this model was also the best-fitting model in their previous work on cabozantinib for the treatment of second line RCC (response to clarification question A24), which the ERG assumes refers to NICE TA463.[14]

The CS presents fitted fractional polynomial survival curves (first and second order) for the outcomes of OS and PFS for all three treatments, based on fixed effects (CS Figure 21 to Figure 40, Appendix D1.1). The CS did not supply hazard ratio plots for each fractional polynomial model with credible intervals to allow visual inspection of the time-varying HRs.

These were requested from the company as well as the tabulated HRs for each interval of the follow-up period (clarification question A22). The company were also asked to provide fractional polynomial results based on a random effects model (clarification question A23). The company provided the requested data and these are described and discussed in section 3.3.7 of this report.

3.1.7.7 Bayesian statistical methods used in the Ouwens and fractional polynomials ITCs

The ERG noted that limited details of the Bayesian methods used to run both the Ouwens et al[36] and fractional polynomial models[40] are given in the CS. Details lacking included the prior probability distributions (e.g. vague, informative, non-informative, the rationale for their choice), the likelihood distribution, the number of iterations used for burn in and inferences, and the methods for assessing convergence. The company were requested to provide this information (clarification question A21).

The company reported that in the Ouwens et al[36] method non-informative priors were used for all models, with model parameters estimated using a Markov Chain Monte Carlo Gibbs algorithm in WinBUGS software. For fixed effects models, three parallel chains were run, with 50,000 iterations for burn in and a further 100,000 iterations for inferences. These were increased to 150,000 and 200,000 iterations respectively for the random effects models.

For the fractional polynomial method the choice of prior was also non-informative and a Markov Chain Monte Carlo Gibbs algorithm in WinBUGS software was also used. Three parallel chains were run with 250,000 iterations for burn in and a further 250,000 iterations for inferences. The Gelman-Rubin statistic Rhat was calculated and convergence declared when Rhat < 1.05 for both the Ouwens and fractional polynomial methods. Rhat is a standard model convergence statistic reported in WinBUGS; values close to 1.0 (i.e. <1.05) are considered indicative of convergence.[44] The company did not report whether or not they had conducted sensitivity analyses on choice of prior.

Based on the information provided the ERG considers that the methods used to implement the two ITC methods are appropriate and correspond to the methods specified in the original methodological texts.[36 40]

3.1.7.8 Summary of the ERG's appraisal of the ITC

- The company conducted ITCs to compare cabozantinib against pazopanib given the lack of head-to-head evidence for these two treatments.

- The company's ITC includes two RCTs: CABOSUN (cabozantinib versus sunitinib) and COMPARZ (sunitinib versus pazopanib). CABOSUN was a phase II RCT (n=157 patients) whilst COMPARZ was a larger phase IIIb non-inferiority RCT (n=1110 patients).
- These two trials have some similarities:
    - Treatments were administered continuously until progression of disease, the occurrence of unacceptable toxic effects, or withdrawal of consent; identical dose regimen of sunitinib were used; mean age and gender profile was similar; all patients had clear cell RCC and most patients had metastatic disease.
    - However, there are some important differences: the CABOSUN trial included only patients at intermediate or poor RCC risk, whilst the COMPARZ study included patients with favourable, intermediate and poor risk classifications; Around a third of patients in CABOSUN had bone metastases (a key prognostic factor in RCC) at baseline compared to 18% of patients in COMPARZ; a greater proportion of COMPARZ patients were classified as having the highest cancer performance status. The impact of these differences on the results of the ITC are not discussed in the CS. The ERG considers that they may under-estimate the relative effectiveness of cabozantinib versus pazopanib.
- The ERG considers the methodological quality of the two trials to be adequate overall and the overall risk of bias to be low, though there is risk of bias relating to blinding due to the open-label nature of the trials. The CABOSUN trial has some further limitations (i.e. phase II trial, relatively small sample size; immature OS data).
- Due to the observation that proportional hazards do not hold for all survival outcomes in both trials, the CS used ITC methods that do not assume proportionality in hazards. These were the ITC of parametric survival curves using methodology developed by Ouwens et al,[36] and the use of parametric models with fractional polynomial distributions using methodology developed by Jansen et al.[40] The Ouwens et al method provides survival estimates for a family of parametric distributions (Weibull, log-logistic, log-normal, Gompertz, exponential). The fractional polynomial method provides survival estimates for first order and second order models from a set of powers (five models for each order, 10 models in total). Both of these methods provide survival effect estimates that are used in the company's economic model.
- The Ouwens and fractional polynomial methods appear to have been implemented adequately in accordance with the original publications,[36 40] and the ERG considers

that both are suitable for use for the indirect comparison of treatments in this appraisal. However, the results of both methods may be biased by the aforementioned differences between the two trials in RCC risk factors and other variables.

- The results of the ITC based on these methods are described later in this report (section 3.3.7 and their suitability for use in the economic model to inform cost-effectiveness estimates are discussed in section 4.3.4).

3.2 Summary statement of company's approach to evidence synthesis

The ERG's assessment of the company's approach to the evidence synthesis is summarised in Table 9.

Table 9 Quality assessment (CRD criteria) of CS review

CRD Quality Item	ERG response
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Yes. Eligibility criteria are tabulated (CS Table 4) and are generally appropriate, but with the following minor inconsistencies: • The tabulated inclusion criteria for interventions and comparators are broader than those finally applied to identify eligible studies. Final eligibility criteria were stated as "only treatments in the NICE scope of the appraisal" (footnote in CS Table 4) and "only publications related to cabozantinib, sunitinib and pazopanib were included in the final selection" (CS section B.2.1). Response rates are listed in the company's decision problem but are not specified as an outcome in the inclusion criteria.
2. Is there evidence of a substantial effort to search for all relevant research?	Yes. Systematic literature searches were based on a search conducted by the manufacturer of pazopanib (1980-2009) which the company updated to June 2017 and widened to include cabozantinib and tivozanib (CS section B.2.1). The overall search was comprehensive and wider than the NICE scope, although the company did not systematically search specific conferences. The ERG ran updated searches to March 2018 and did not identify any further relevant studies.
3. Is the validity of included studies adequately assessed?	Yes. The company assessed the risk of bias in the CABOSUN trial, as well as in the COMPARZ trial that was included in the company's ITC analysis (Table 15 in CS Appendix D).
4. Is sufficient detail of the individual studies presented?	Partly. The study methods (CS Tables 6-9 and 11), baseline characteristics of the participants (CS Table 10), and participant flow (Figure 52 in CS Appendix D) are clearly reported for the CABOSUN trial. Baseline characteristics of the COMPARZ trial included in the company's ITC analysis are also clearly reported (Table 11 in CS Appendix D), but limited detail on the COMPARZ trial methods is provided (Table 10 in CS Appendix D) and patient flow is not reported.
5. Are the primary studies summarised appropriately?	Yes. Results from the CABOSUN trial are clearly summarised for all clinical effectiveness outcomes (CS section B.2.6). Results from the COMPARZ trial are

	summarised in CS Tables 20 for those outcomes relevant to the ITC (PFS and OS).

The company's evidence synthesis is generally well structured and clearly reported. The company's search for clinical effectiveness studies identified a broader range of interventions and comparators than those specified in the NICE scope. The company subsequently restricted the intervention and comparators at the eligibility screening step to be consistent with the scope. Overall, the company's evidence synthesis is consistent with their decision problem and the NICE scope, with the exception that HRQoL, which is an outcome specified in the scope, was not reported in the CABOSUN trial. HRQoL is therefore not included in the company's decision problem and not reported in the clinical effectiveness synthesis in the CS; a separate systematic review of utility studies was conducted to inform the company's economic analysis (CS section B.3.4).

3.3 Summary of submitted evidence

In the following sub-sections we summarise the results of the CABOSUN trial, based on data reported in the CS for the IRC assessment to determine progression and response and the FDA recommended censoring rules, for the most recent data cut-off date available. These are based on data in the CSR[27] and a 2017 European Society for Medical Oncology (EMSO) conference poster.[30] We do not present results from the earlier trial journal publication,[24] as these are based on an earlier data cut-off date (April 2016); are based on investigator rather than IRC assessment; use non-FDA censoring rules (and were not used in the company's regulatory submission); and are not used in the economic model.

3.3.1 Summary of results for progression free survival (PFS)

Figure 6 shows the Kaplan-Meier PFS curves, based on the September 2016 data cut-off.[27] [30] At a median follow-up of 25 months, median PFS was 8.6 months (95% CI 6.8, 14.0) for cabozantinib and 5.3 months (95% CI 3.0, 8.2) for sunitinib (p=0.0008). The median difference was 3.3 months. The HR, stratified by IMDC risk category and bone metastases, was 0.48 (95% CI 0.31, 0.74). As can be seen from Figure 6, the survival curves appear parallel from month three onwards, implying proportional hazards. The majority of events recorded were for documented disease progression: 40 (51%) in the cabozantinib group; 43 (55%) in the sunitinib group. The remaining patients were censored: 36 (46%) in the cabozantinib group; 29 (37%) in the sunitinib group (CS Table 12). PFS at 12 months (% event free) was 43.1 and 21.1 in the cabozantinib and sunitinib groups, respectively.

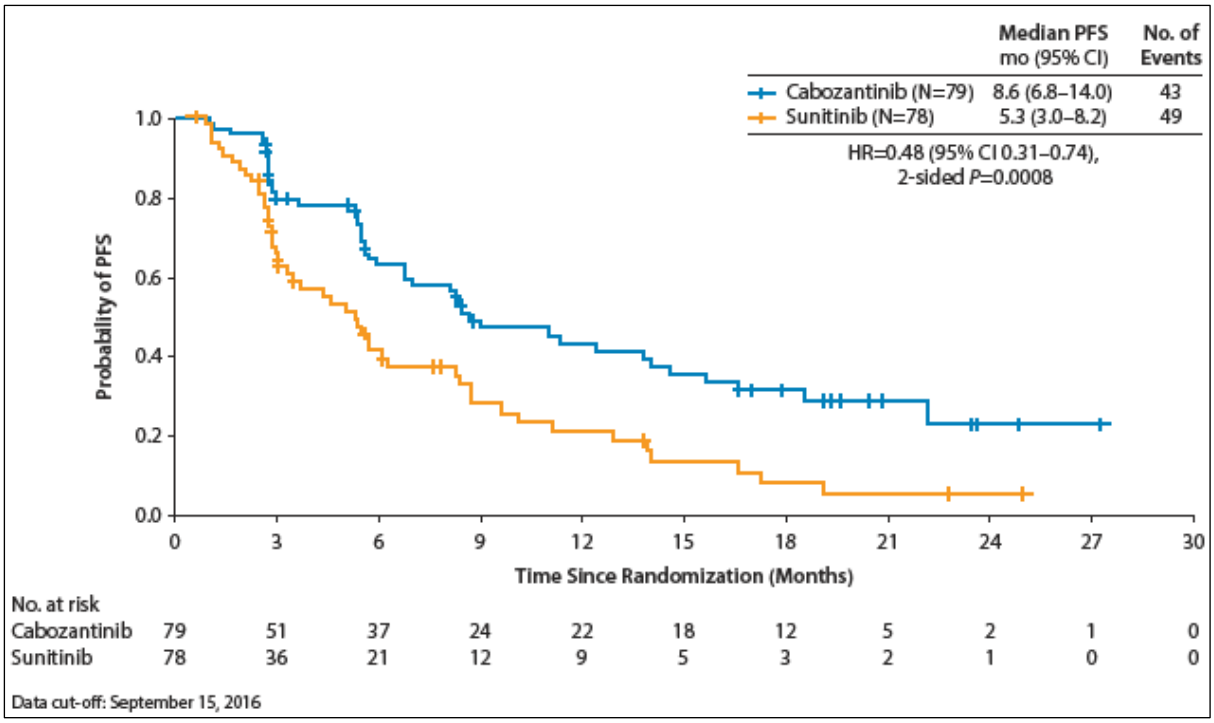


Figure 6 Kaplan-Meier PFS curves (IRC, ITT population. Reproduced from CS Figure 5)

As mentioned earlier, PFS was the primary outcome of the CABOSUN trial. However, the ERG notes that this would have been under-powered statistically, since a 2-sided test, as used in the IRC-based analysis conducted for the submission to the regulator (and also used in the CS) has less statistical power than a 1-sided test (used in the original trial analysis), and the 92 events is fewer than the planned target of 123 events (see section 3.1.6 of this report for more information on the statistical procedures used in the trial).

3.3.2 Summary of results for overall survival (OS)

Overall survival was a secondary outcome of the CABOSUN trial. Figure 7 shows the Kaplan-Meier OS curves, based on the January 2017 data cut-off. At a median follow-up of 28.9 months, the median OS was 30.3 months (95% CI 14.6, not estimable) in the cabozantinib arm versus 21.0 months (95% CI 16.3, 27.0) in the sunitinib arm. The median difference was 9.3 months. The CS notes that the data were immature at this data cut-off and there was a notable degree of censoring around the median estimates (censoring due to no event as of the cut-off date – 52% and 42% of patients in the cabozantinib and sunitinib groups, respectively). Thus the OS data should be interpreted with caution. The HR, stratified by IMDC risk category and bone metastases, was 0.74 (95% CI 0.47, 1.14) p=0.1700. As can be seen from Figure 7, the survival curves cross at around month 14 before crossing again and then separating at around month 21 for the rest of the follow-up period. Proportional hazards do not therefore appear to hold.

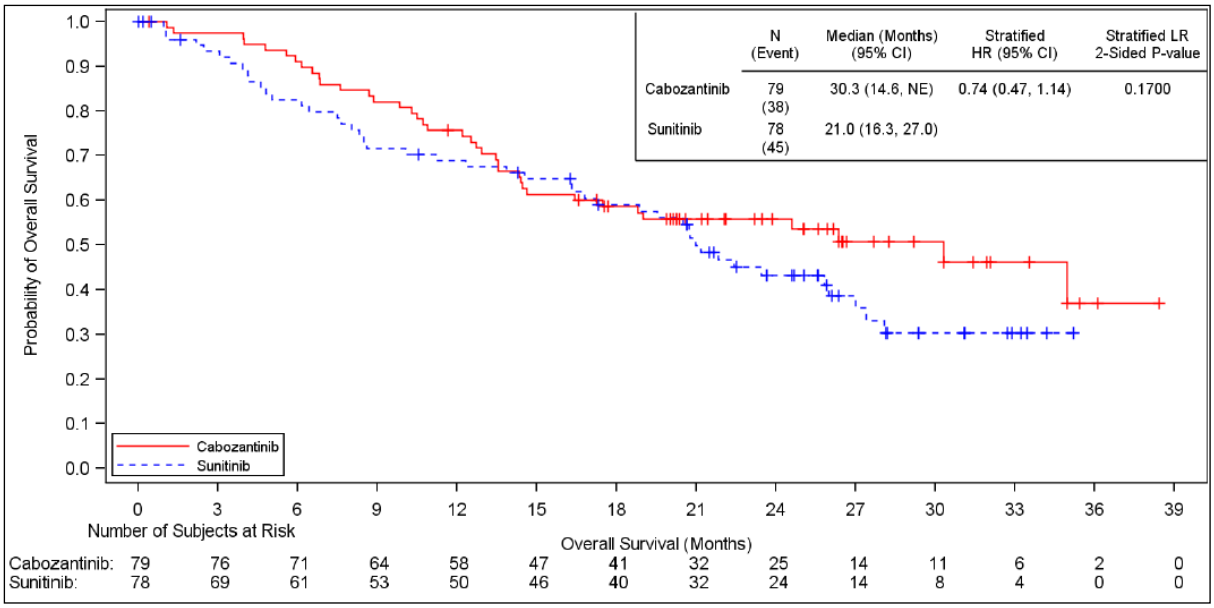


Figure 7 Kaplan-Meier plot of OS (13th January 2017 data cut-off, ITT population. Reproduced from CS figure 6)

The percentage of patients event-free at 30 months was 50.7% and 30.3% for cabozantinib and sunitinib, respectively (CS Table 13). The CS states that these OS data are used to inform the economic model.

Figure 8 illustrates an updated OS analysis, at the data cut-off of 1st July 2017 (thus around six months after the above OS analysis; median follow-up not reported), presented at the EMSO conference.[30] As can be seen, the median difference in OS between the treatments is 5.5 months, favouring cabozantinib. However, the confidence intervals around the OS estimates are wide and the confidence interval around the HR crosses 1, indicating a non-statistically significant difference. Data from this cut-off do not appear to have been used in the economic model, and it is not stated in the CS why data from the earlier OS data cut-off (January 2017) were used in preference.

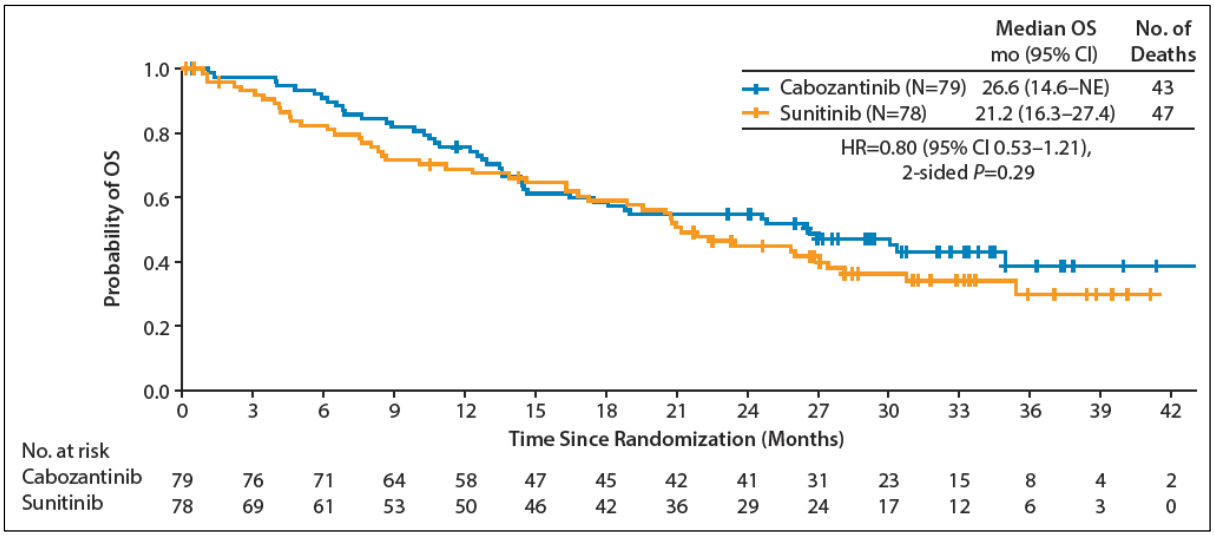


Figure 8 Kaplan-Meier plot of OS (July 2017 data cut-off, ITT population. Reproduced from CS figure 7)

The ERG notes that the OS estimates presented will have been influenced by subsequent anticancer treatments that trial participants received on discontinuation of the study treatment (systemic non-radiation anticancer therapy was received by 57%-58% of patients) (see section 3.3.4 below). The CS does not discuss the impact of these treatments on the OS estimates, or make any adjustments to the OS estimates in the economic model. The impact of subsequent anti-cancer treatments on OS is therefore uncertain. (NB. The ERG adjusts the costs to reflect different assumptions about subsequent anticancer treatments in a scenario analysis – see section 4.4).

In summary, the ERG urges caution in the interpretation of the OS results from this study as the data are immature, the survival curves cross each other indicating non-proportional hazards, the study was not statistically powered for OS, and the uncertain influence of subsequent anticancer treatments received by a large proportion of patients in both study groups.

3.3.3 Summary of results for tumour response

CS Table 14 presents tumour response data based on IRC assessment in the ITT population (data cut-off September 2016). As stated earlier, this outcome is not used to inform the economic model. The objective response rate (ORR) was 20% (95% CI 12.0%, 30.8%) in the cabozantinib arm, compared to 9% (95% CI 3.7%, 17.6%) in the sunitinib arm, classed as a 'confirmed partial response'. The difference between groups in ORR was 11.3% (95% CI, 0.4 22.2%; p=0.0406). There were no confirmed complete responders in either study group. The proportion of patients with stable disease was higher in the

cabozantinib group (n=43; 54%) than the sunitinib group (n=30; 38%). The proportion of patients with progressive disease was lower in the cabozantinib group (n=14; 18%) than in the sunitinib group (n=23; 29%). The CS also reports the disease control rate, defined as complete response + partial response + stable disease at 75% and 47% in the cabozantinib and sunitinib groups, respectively. In actuality, this rate only reflects partial response and stable disease as there were no complete responses in the trial.

CS Table 15 reports tumour response results for the three sets analyses (Investigator-assessed, Alliance censoring rules, April 2016 cut-off; Investigator-assessed, FDA censoring rules, September 2016 cut-off ; IRC-assessed, FDA censoring rules, September 2016 cut-off). This permits side-by-side comparison of the results at different cut-offs/tumour assessment/censoring rules. We have not reproduced this table here, but in summary we note that the ORR for both study groups is lower (and the between group difference is smaller) for the IRC assessment using FDA censoring rules (i.e. the data used in the company's regulatory assessment).

3.3.4 Subsequent anti-cancer treatment

Although crossover was not permitted in the trial, patients could receive subsequent non-protocol treatments upon discontinuation of the study treatment (e.g. on disease progression). Whilst not an outcome in the scope of the appraisal, the CS reports details of subsequent treatments given (CS Table 16 and Appendix L). The proportion of patients receiving any systemic non-radiation anti-cancer was similar: 45 (57%) and 45 (58%) in the cabozantinib and sunitinib groups, respectively. The median time to first systemic non-radiation anti-cancer therapy was 196 (range 56, 877) days and 147 (range 4, 725 days) in the cabozantinib and sunitinib arms, respectively. Just under half of all patients received a VEGFR-targeted TKI drug as a subsequent treatment (axitinib, pazopanib, sunitinib, cabozantinib, or sorafenib). The ERG notes that sunitinib and sorafenib have not been recommended by NICE for second-line RCC treatment. Around 14% of patients overall received an anti-PD-1/PD-L 1 targeting agent as subsequent therapy, including nivolumab. Other systematic therapies used as subsequent treatment included temsirolimus, everolimus and bevacizumab.

The company were asked to clarify the number of patients who received each subsequent line of systemic anticancer therapy (e.g. second, third, fourth line etc) (clarification question A19). The company clarified that only first non-protocol treatments and concomitant medications were captured in the case report forms. Thus it appears that the data provided on subsequent treatments refer to second-line treatment only. However, in contradiction, the

ERG notes that Appendix L Table 50 states to include all reported subsequent anti-cancer treatments (including “first subsequent treatment and any further treatments reported”).

3.3.5 Summary of Health related quality of life (HRQoL)

As stated earlier, HRQoL was not measured in the CABOSUN trial. For details of the company’s HRQoL utility estimates in the economic model see section 4.3.5 of this report.

3.3.6 Sub-group analyses results

CS Appendix E Figure 53 provides a forest plot showing pre-specified subgroup analyses for the outcome of PFS as determined by IRC assessment (see section 3.1.6 earlier in this report for details of the statistical procedures used in the subgroup analyses). The CS comments that there was a consistently favourable effect for cabozantinib compared with sunitinib in larger subgroups (≥ 20 patients). There was a favourable effect for cabozantinib compared to sunitinib in the following subgroups: age (<65 years ≥ 65 years); sex (male); race (white); baseline ECOG status (0); bone metastases (yes/no); RCC risk factors (intermediate/poor); and MET status (positive). Confidence intervals were wide and included 1 for some the smaller subgroups (e.g. race group ‘other’). Of note, the PFS HR was more favourable for the poor RCC risk subgroup (0.31, 95% CI 0.11, 0.92) than the intermediate risk group (0.52, 95% CI 0.32, 0.82), though this is based on a very small sample of patients (15 poor risk patients in each group).

The CS did not present subgroup analyses for the outcome of OS, but supplied them on request (clarification question A9) in a table, with no commentary or interpretation. The results appear to be based on the January 2017 data cut-off. Overall, the results were consistent with the overall population analysis results, with OS more favourable in the cabozantinib group than the sunitinib group. However, in most subgroups the confidence intervals included 1 (as in the overall population analysis). Tests for treatment by subgroup interaction yielded non statistically significant p values except for MET status.

The CS does not present subgroup analyses for the outcome of tumour response. The ERG identified a conference abstract presented at the American Society of Clinical Oncology (ASCO) in February 2018 which reported subgroup analyses of the CABOSUN trial for the PFS and ORR outcomes.[26] The PFS results are the same as those reported in the CS (summarised above). Odds ratios for ORR are given for the following subgroups: IMDC risk group, bone metastases, age, sex, baseline ECOG and MET status. No confidence intervals around the odds ratios are given, or any other descriptive statistical information. The data

show odds ratios greater than 1 for all subgroups, and the abstract states that odds ratios favours cabozantinib over sunitinib. The ERG interprets this as a higher odds of achieving a confirmed partial response with cabozantinib (as was the case in the overall study population – see section 3.3.3 above).

The ERG urges caution in the interpretation of subgroup analyses as they are not statistically powered to detect a difference between treatments, and some of the subgroups are quite small leading to uncertainty in effects. In particular, the OS subgroup results require caution for the aforementioned limitations of immature data, non-proportional hazards and uncertain influence of subsequent anticancer treatments. To reiterate, the scope of this appraisal does not specify any relevant subgroups for assessment.

3.3.7 Indirect treatment comparison results

3.3.7.1 ITC results: comparison of parametric survival curves

The CS reports the results of the ITC as fitted survival curves for the outcomes of OS and PFS for all three treatments (cabozantinib, sunitinib, pazopanib), based on fixed effects and random effects, for each of the five parametric distributions generated by the Ouwens et al method.[36] For each of the analyses cabozantinib had a higher survival estimate than sunitinib or pazopanib.

It is not practical to show all of the graphs here, but to illustrate, Figure 9 below shows the PFS fitted curves based on the log-normal model which was selected by the company as the most appropriate model to inform the economic model (CS Table 33). Figure 10 below reports the OS fitted curves based on the exponential model as this was selected by the company as the most appropriate model to inform the economic model (CS Table 33). The sunitinib and pazopanib curves were similar to each other in shape and position, indicating similar effectiveness, as was the case in all of the other fitted parametric survival models (CS appendix D1.1).

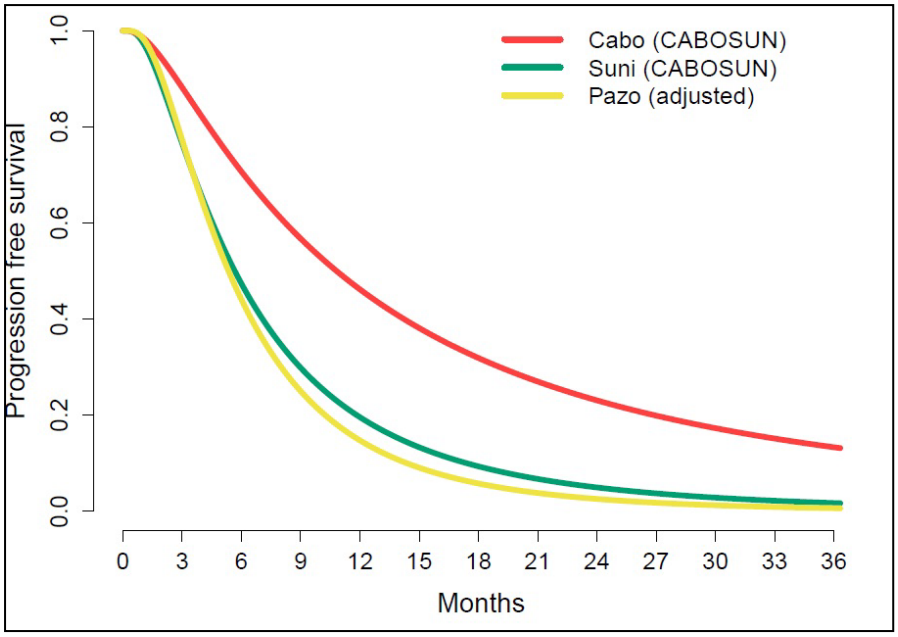


Figure 9 PFS ITC results, Ouwens model, log-normal distribution, fixed effect (reproduced from CS Appendix D1.1 Figure 14).

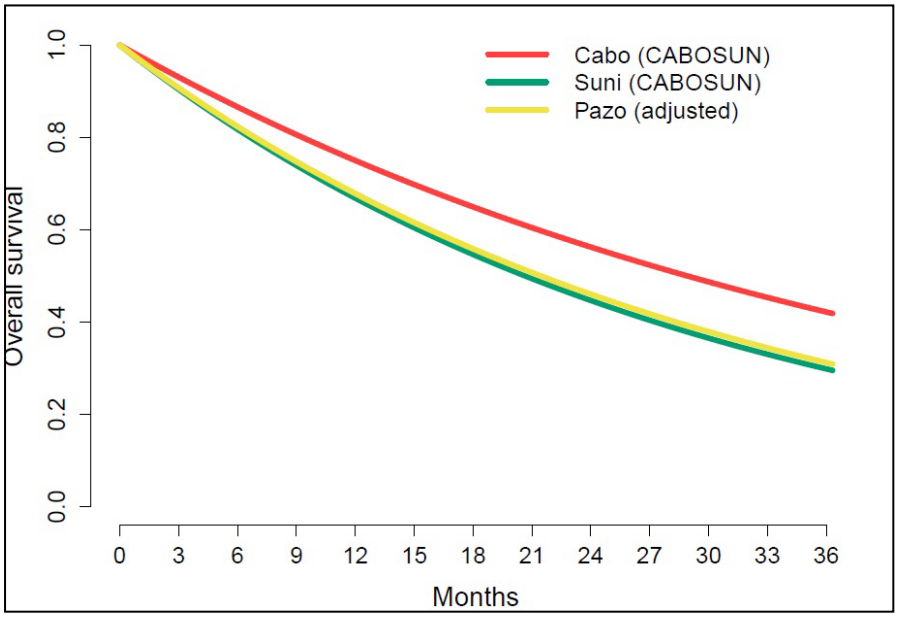


Figure 10 OS ITC results, Ouwens model, exponential distribution, fixed effect (reproduced from CS Appendix D1.1 Figure 1).

3.3.7.2 ITC results: fractional polynomials

The CS presents fitted fractional polynomial survival curves for the outcomes of OS and PFS for all three treatments, based on fixed effects for first and second order models (CS Figure 21 to Figure 40, Appendix D1.1). The CS did not supply hazard ratio plots with credible intervals for each fractional polynomial model to allow visual inspection of the time-varying HR curves. These were requested from the company as were the tabulated HRs for each

time interval of the follow-up period (clarification question A22). These were provided by the company with the time period split into monthly intervals.

It is not practical to show all of the graphs here, but for illustration, Figure 11 shows the PFS hazard ratio plot for the company's best-fitting fractional polynomial model (second order P1=-1 and P2=-1) which informed the economic model (the tabulated HRs for these plots are reported in Table 7 and Table 8 of the company's clarification question A22 response). As can be seen:

- The HR for pazopanib peaks at month four [redacted] and declines slightly during the rest of the follow-up period. The HR for sunitinib peaks at month six [redacted] and declines slightly during the remainder of the follow-up period.
- The credible intervals increase over the follow-up period, with the upper bound increasing to include 1 after month 19 for pazopanib, and after month 11 for sunitinib.
- The ERG notes that the time-varying PFS HRs for cabozantinib versus sunitinib generated by this fractional polynomial model ITC compare broadly with the constant HR reported in the CABOSUN trial (0.48 (95% CI 0.31, 0.74), though there is greater uncertainty in the fractional polynomial model as evident from the wide credible intervals which include 1 for a large proportion of the follow-up period.

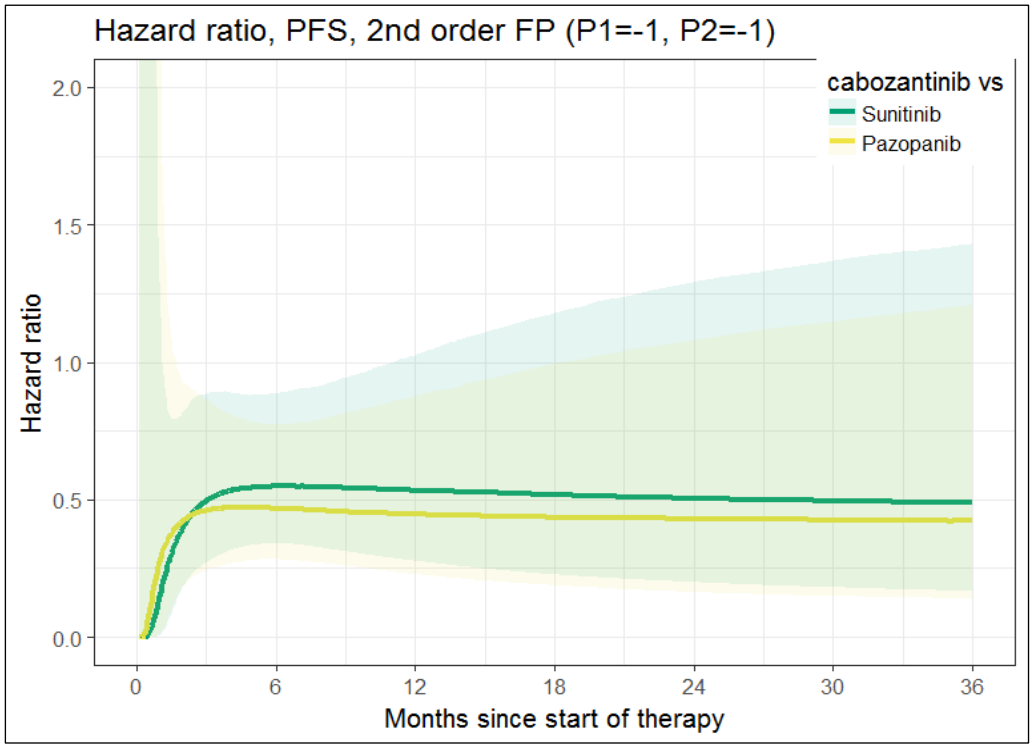


Figure 11 Hazard ratio plot, PFS; fractional polynomial 2nd order (p1=-1, p2=-1), fixed effect (reproduced from company clarification question response A22 CS figure 28)

Figure 12 shows the OS hazard ratio plot for the company's best-fitting fractional polynomial model (second order P1=-1 and P2=-1) used to inform the economic model (the tabulated HRs for these plots are reported in Table 3 and Table 4 of the company's clarification question response).

- The HR for pazopanib starts to peak at month nine, and declines slightly after month 19 [redacted] The HR for sunitinib begins to plateau at month 13 and peaks at month 30 where it remains for the rest of the follow-up period [redacted].
- The credible intervals widen during the course of the follow-up period, and include 1 at all time points.
- The ERG notes that the time-varying OS HRs for cabozantinib versus sunitinib generated by the fractional polynomials ITC compare broadly with the constant OS HR reported in the CABOSUN trial (0.80 (95% CI 0.53, 1.21), though there was greater uncertainty in the fractional polynomial model as evident from the wide credible intervals.

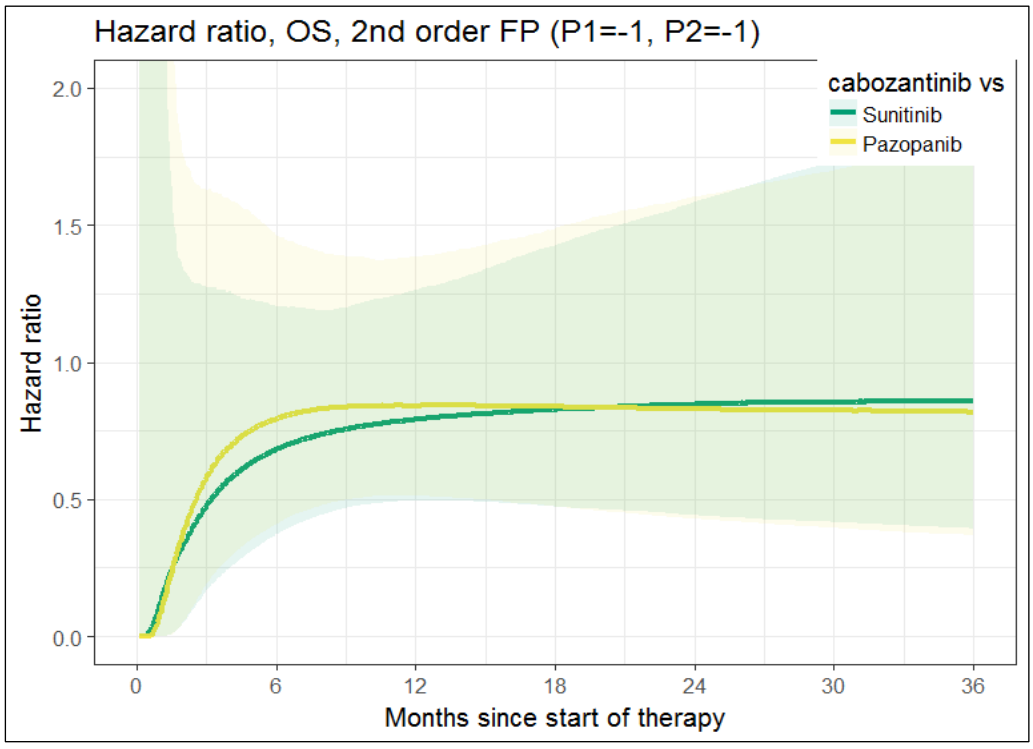


Figure 12 Hazard ratio plot, OS; fractional polynomial 2nd order (p1=-1, p2=-1), fixed effect (reproduced from company clarification question response A22 Figure 18)

The ERG has reviewed the results of the other fractional polynomial models (as supplied in response to clarification question A22, Figures 11 to 30). Our general observation is that, across the different models, the time-varying HR curves for cabozantinib versus sunitinib

and cabozantinib versus pazopanib have a similar shape to each other. Cabozantinib is of superior effectiveness when compared with both sunitinib and with pazopanib, with little difference between the results of each pairwise comparison.

Appendix 9.4 of this report provides additional ITC results:

- A summary of the results of the other (i.e. the non-best fitting) fixed effect fractional polynomial models.
- A comparison of the results of random effects and the fixed effect fractional polynomial models.
- A comparison of the results from the ITC using the wider evidence network with the restricted evidence network.

3.3.8 Summary of adverse events

CS section B.2.10 summarises adverse reactions recorded in the CABOSUN trial. Table 10 below summarises the incidence of adverse events. As mentioned earlier (section 3.1.6), adverse events were assessed in the safety analysis population, defined as all patients who received any treatment with cabozantinib or sunitinib. The safety population comprises 78/79 (99%) of patients randomised to the cabozantinib group, and 72/78 (92%) of patients randomised to sunitinib. Thus, there was a slight imbalance in the size of the study groups in this population. Adverse events were described as solicited (expected per the protocol and presence/absence and severity solicited at baseline and for each treatment cycle), and unsolicited (other adverse events not expected). The CS states that the safety data reported are taken from the CSR[27] and may differ from the trial journal publication[24] due to regulatory reporting requirements. The ERG notes that the safety data do indeed differ between these two publications, and the data in the CS (i.e. based on the CSR) should therefore be considered definitive. These data are summarised below.

The duration of treatment exposure was longer in the cabozantinib arm compared with the sunitinib arm (median: 6.5 months versus 3.1 months). Dose reductions were reported to be frequent with both treatments: (46% of cabozantinib patients; 35% of sunitinib patients) as were dose interruptions (73% and 71% respectively).

The percentage of patients with at least one treatment-related adverse events was similar between the two study groups (95%-97%). Grade 3 or 4 adverse events were reported in a similar percentage of patients in the study groups (60%-63%), as were serious adverse events (49%-51%) and treated-related serious adverse events (36%). Discontinuations of

study drug due to adverse events was also similar between study groups (21%-22%). The percentage of patients dying up to 30 days after last dose of study treatment was higher in the sunitinib group compared to the cabozantinib group (11% versus 5.1%, respectively), as was the case for death > 30 days after last dose of study treatment (49% versus 44%, respectively).

Table 10 Summary of AE incidence (safety population) (reproduced from CS Table 25)

	Cabozantinib N = 78 n (%)	**Sunitinib N = 72 n (%)**
AE	75 (96)	71 (99)
Related AE	74 (95)	70 (97)
Worst AE, grade 3 or 4	53 (68)	47 (65)
Worst related AE, grade 3 or 4	47 (60)	45 (63)
Grade 5 AE up to 30 days after last dose of study treatment[a]	3 (3.8)	6 (8.3)
Grade 5 AE > 30 days after last dose of study treatment	1 (1.3)	3 (4.2)
Related grade 5 AE at any time	2 (2.6)	4 (5.6)
Serious AE	38 (49)	37 (51)
Related serious AE[b]	28 (36)	26 (36)
Deaths	38 (49)	43 (60)
Death up to 30 days after last dose of study treatment	4 (5.1)	8 (11)
Death > 30 days after last dose of study treatment	34 (44)	35 (49)
Discontinuation of study due to AE[c]	21%	22%

[a] Grade 5 AEs were not reported for 3 subjects (1 cabozantinib, 2 sunitinib) who died < 30 days after the last dose of study treatment; [b] grade 1 or 2 SAEs that did not entail hospitalisation ≥ 24 h were not recorded in the clinical database; [c] based on patient disposition, not excluding events of disease progression, only % reported. 'Unsolicited' grade 1 and 2 events not related to study treatment were not collected.
AE, adverse event.

CS Table 26 reports the percentage of patients experiencing specific treatment-related adverse events. The incidence of specific events varied between the study groups. Common adverse events (of any grade) in the cabozantinib arm were diarrhoea (72%), fatigue (62%), aspartate aminotransferase increased (60%), hypertension (56%), alanine aminotransferase increased (54%), decreased appetite (45%) and palmar-plantar erythrodysaesthesia syndrome (42%). Of these, all except decreased appetite was a solicited adverse event. In the sunitinib group common adverse events (of any grade) included fatigue (67%), platelet count decreased (58%), diarrhoea (49%), anaemia (44%) hypertension (38%), nausea (36%) and neutrophil count decreased (35%). Of these, all except anaemia and nausea were a solicited adverse event.

Common grade 3/4 adverse events in the cabozantinib arm included hypertension (22%), diarrhoea (9%), hypophosphataemia (9%), palmar-plantar erythrodysaesthesia syndrome (7.7%), fatigue (5.1%), decreased appetite (5.1%), and stomatitis (5.1%). In the sunitinib arm common grade 3/4 adverse events included hypertension (18%), fatigue (17%), platelet count decreased (11%), diarrhoea (8.3%), and hypophosphataemia (6.9%).

Similar specific adverse events (of any grade, and grade 3/4) were common in both treatment groups, though the percentage of patients experiencing them varied between the groups.

4 COST EFFECTIVENESS

4.1 Overview of company's economic evaluation

The company's submission to NICE includes:

i) A review of published economic evaluations of cabozantinib compared with sunitinib and pazopanib for patients with untreated locally advanced or metastatic RCC.

ii) A report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of cabozantinib is compared with sunitinib and pazopanib for treatment-naïve patients with advanced RCC.

4.2 Company's review of published economic evaluations

The company conducted a systematic search of the literature to identify economic evaluations with cabozantinib or its comparators sunitinib, pazopanib in untreated advanced RCC. Details of the review methods are reported in CS Appendix G. It included cost-effectiveness studies of selected first line treatment options (sunitinib, pazopanib, interferon-alfa, interleukin-2, bevacizumab + interferon-alfa, temsirolimus, sorafenib and tivozanib) for patients with advanced/metastatic, previously untreated RCC. The search was not restricted by timeframe, language (other than English, German, French, Spanish and Italian were excluded) or countries (other than European countries, Australia, Canada were excluded).

The inclusion criteria state that full-text publications, conference abstracts and reports were included while letters, editorials, notes, and historical articles were excluded. The search identified 804 papers, which were assessed against predefined inclusion/exclusion criteria (Appendix G, Table 17). One cost-effectiveness study was excluded due to language barriers (Czech Republic). A total of 35 studies were excluded due to a focus on different countries. Table 21 (CS, Appendix G) presents the references excluded due to country.

Of the 23 studies included in CS (Table 22, CS Appendix G), 9 were critically appraised using the Drummond and Jefferson checklist (1996). Of the remaining studies, 9 were only available as conference abstracts or posters and 5 were technology appraisals published by technology assessment agencies (4 of them by NICE). Summary results of the critical appraisal are presented in Tables 23 and 24 (CS, Appendix G). Table 25 shows the studies that were not assessed with reasons (CS, Appendix G).

Table 29 (CS B.3.1) summarises the methods and results of seven studies that were conducted from an English, Welsh or British perspective. The company concluded that as

none of these studies included cabozantinib, they are not directly relevant to this appraisal. The ERG agrees with this conclusion.

4.3 Company's submitted economic evaluation

4.3.1 NICE reference case

Table 11 shows that the company's economic evaluation adheres to NICE's reference case requirements.

Table 11 NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per NICE scope	Yes	Although PFS and OS curves from ITC also include patients with favourable risk status.
Comparator: As listed in NICE scope	Yes	
Perspective on costs: NHS and PSS	Yes	
Evidence on resource use and costs: Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	Yes	
Type of economic evaluation: Cost utility analysis with fully incremental analysis	Yes	
Synthesis of evidence on outcomes: Based on a systematic review	Yes	
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	20 years in base case (10 in scenario analysis)
Measuring and valuing health effects: Health effect should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life.	Yes	HRQoL not collected in CABOSUN. EQ-5D estimates from published sources used
Source of data for measurement of health-related quality of life: Reported directly by patients and/or carers.	Yes	
Source of preference data: Representative sample of the UK population	Yes	
Equity considerations: An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	Yes	
Discount rate: 3.5% per year for costs and health effects	Yes	

4.3.2 Modelled decision problem

The model broadly reflects the decision problem in the scope, but with some uncertainties.

Population: The model uses a cohort with an initial age (62.8 years) and gender mix (78% male) similar to that in the CABOSUN and COMPARZ populations (Table 12). The ERG has been advised that in practice, patients starting first-line treatment for advanced RCC are often older than trial participants. We explore the impact of age on cost-effectiveness through scenario analysis to assess the applicability of the results.

Table 12 Population characteristics in the model and comparative statistics

Baseline characteristics	Model	CABOSUN[24]	COMPARZ[34]	IMDC database cohort[7]
Age (years)	62.8	Median 63	Median 61/62	55% 60+
Male	78.3%	78%	73%	74%
Favourable risk	Not explicit	0%	25%	18%
Intermediate risk		81%	55%	52%
Poor risk		19%	18%	30%

The distribution by IMDC risk group is not specified in the model but is set implicitly by the sources of effectiveness evidence. As discussed in section 3.1.7.1 above, there is an important question over how well the ITC model reflects the scope population because of the inclusion of favourable risk patients in COMPARZ. We consider the implications of this potential source of bias in relation to the choice of PFS and OS effectiveness parameters for the model.

Subgroups: The CS does not present cost-effectiveness for any patient subgroups (CS B.3.9). This is in accordance with the scope, and the company notes that CABOSUN showed consistent results across a range of subgroups (CS Appendix E). The ERG agrees that investigation of cost-effectiveness for subgroups is not warranted given available evidence, but we urge caution over interpretation of the subgroup analyses of trial data as these are not powered to detect a difference (section 3.3.6 above).

Intervention and comparators: The model compares the cost-effectiveness of first-line cabozantinib in comparison with sunitinib and pazopanib, as specified in the scope (CS B.3.2). NICE guidance recommending tivozanib in this indication[22] was published after finalisation of the scope and submission of the CS, so is not included as a comparator in the company model. We do not consider this further.

Outcomes: The model reflects the outcomes specified in the scope. Quality of life data was not collected in CABOSUN, so utilities for health states and adverse events are based on published sources for patients receiving other treatments (CS B.1.1). We discuss the appropriateness of utility sources in section 4.3.5 below.

4.3.3 Model structure and assumptions

The model structure is described in CS B.3.2 and illustrated in Figure 12, reproduced in Figure 13 below. It is a health state transition (Markov type) model, containing three mutually-exclusive health states: progression free (PF); progressed disease (PD) and death. Patients start in the PF state, following initiation of one of the included treatments at first-line: cabozantinib, sunitinib or pazopanib. At disease progression, patients transition to the PD state, which is considered irreversible, so patients cannot return from PD to PFS. Patients in PF and PD states may die from cancer or other causes.

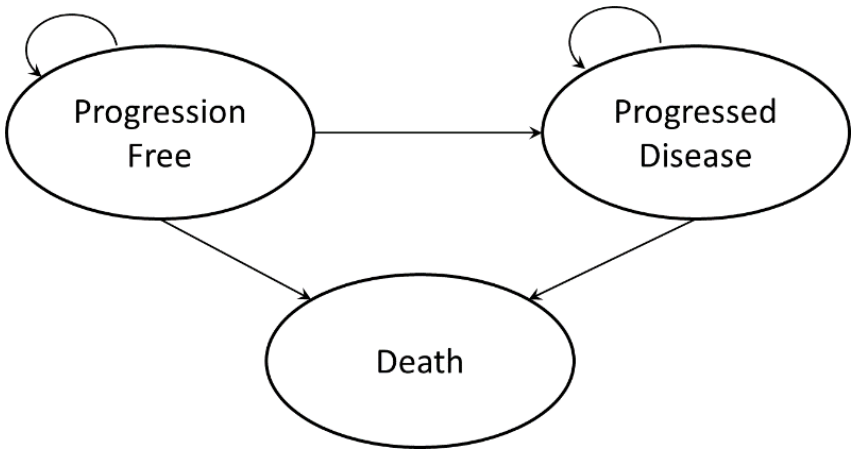


Figure 13 Structure of economic model (reproduced from CS B.3.2 Figure 12)

Alongside the health state transition model, proportions of patients on targeted treatments are estimated as illustrated in Figure 14. Patients enter the PF state on first-line treatment but may stop at any time due to adverse effects or when their disease progresses. Most patients then progress to treatment with one of 10 drugs included in the company's base case after a fixed period of waiting (8 weeks). The duration of second-line treatment is defined for each drug, after which patients are assumed to receive supportive care until death.

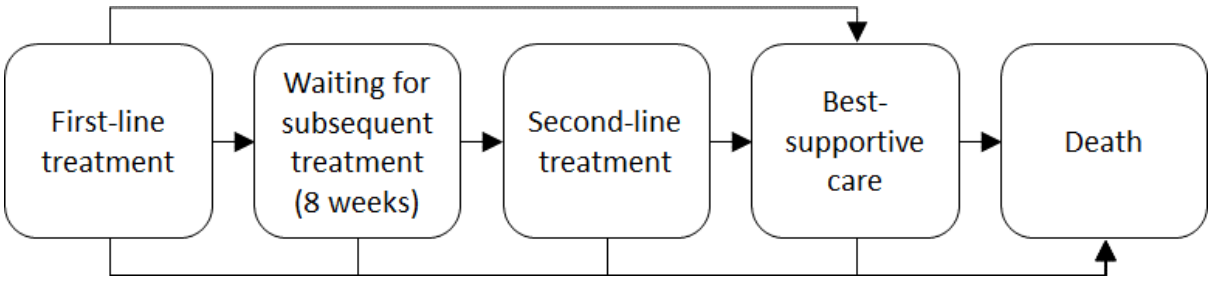


Figure 14 Treatment transition model

The distribution of the cohort between the health states at each time point is estimated using a **partitioned survival approach**, based on PFS and OS curves for the treatment arm:

- **Death**: The proportion of patients alive at each time point is taken from the OS curve. Hence the proportions of the cohort who have died are calculated.
- **PF**: The proportion of patients who are progression free is the minimum of the PFS curve and the OS curve at each time point.
- **PD**: The proportion of patients in the PD state is calculated as the residual (if any) of the cohort who are not dead and not progression free.

Similarly, the distribution of the cohort by treatment status is defined by a Time to Discontinuation (TTD) curve for first-line treatment, a waiting time of 8 weeks between first and second line and fixed treatment durations for the second-line drugs, in addition to PFS and OS curves:

- **First-line treatment**: Calculated from the minimum of the PFS and TTD curves.
- **Waiting for second-line**: The proportion of patients that start waiting in each cycle is calculated based on the proportions who are alive and end first-line treatment. The number of patients waiting is then accumulated over 8 weeks.
- **Second-line treatment**: The proportion of patients emerging alive from the waiting period is calculated and distributed between the 10 active second-line treatments and best supportive care. The time that patients spend on second-line treatment is defined by fixed treatment durations, again adjusting for any deaths within this time.
- **Best supportive care**: Patients who survive the period of second-line treatment enter the best supportive care state, where they remain until they die or the end of the time horizon.

The three-state PF/PD/death model is commonly used in cancer economic evaluations and has been used for previous NICE appraisals for untreated advanced RCC. There is some

controversy over the partitioned survival approach, however, because further assumptions are needed to estimate transition probabilities from survival curves. In this case, the submitted model assumes that the mortality rate is the same pre and post disease progression. This is unlikely but does not affect QALY or health state estimates, which are calculated from the numbers of patients in the three health states at each time point, rather than from the numbers of transitions. The model also assumes the same mortality rate for patients before and after discontinuation of first-line treatment. This does not affect the estimated duration or cost of first-line treatment, which is based on the fitted TTD curve (or PFS if lower). However, it does affect the modelled cost of second-line treatments. If the mortality rate is higher after first-line treatment than before, which seems likely, the model will tend to over-estimate the average duration and cost of second-line therapy.

Other key features and assumptions of the model are listed below:

- **Cycle length**: 1 week, with half cycle correction.
- **Time horizon**: 20 years in base case (with 10 years in scenario analysis).
- **Duration of treatment effects**: based on extrapolation of PFS and OS curves fitted to trial data, assuming persistence of effects over the time horizon.
- **Adverse events**: For each first-line treatment, grade 3 or 4 Treatment Emergent Adverse Events (TEAEs) with an incidence of 5% or more are included in the model. There is no explicit modelling of adverse events related to subsequent treatments.
- **Utility and QALY calculations**: Utility weights for the PF and PD health states are based on published estimates, assumed independent of treatment. Additional disutilities are applied to reflect included TEAEs for first-line treatments – applied as a one-off QALY loss in the first cycle. QALYs are also adjusted for the gender mix and age of the cohort.[45]
- **Health resource use and costs**: The model estimates costs associated with: acquisition and administration of first-line and subsequent treatments, with adjustment for dose intensity and wastage when appropriate; monitoring and disease management in PF and PD states; treatment of included TEAEs for first-line treatments; and end of life care applied in the last cycle before death.
- **Discounting**: 3.5% per year for costs and QALYs
- **Uncertainty**: the model includes macros to conduct: deterministic sensitivity analysis (DSA) with results presented in a tornado diagram; scenario analyses varying selected model assumptions; and probabilistic sensitivity analysis (PSA) producing a cost-effectiveness scatterplot and cost-effectiveness acceptability curve.

The ERG believes that the model structure and partitioned survival approach is appropriate, although we do have some concerns over the following issues:

- It is appropriate to estimate costs and health effects over the patients' whole lifetimes, so we do not disagree per se with the company's use of a 20-year time horizon. Other RCC appraisals have adopted a more conservative time horizon of only 10 years.[12 13 22] In the company's base case model, a relatively small proportion of the modelled cohort survive to 10 or 20 years: about 2% and 0.03% respectively with sunitinib based on CABOSUN survival data. However, we do question the extrapolation of OS and PFS curves from limited trial follow-up over 20 years. This entails strong assumptions about persistence of treatment effects, which may not be realistic. We investigate the impact of the time horizon and different assumptions about persistence of treatment effects in the ERG analysis.
- The model does not include an adjustment for age-related increase in mortality in the general population, as the it relies entirely on the projected OS curves. Given the high rate of mortality for people with advanced RCC, this might not affect results, but we check that the model does not yield counter-intuitive results with longer-surviving RCC patients having lower mortality than members of the general population at the same age.
- The assumption of equal mortality rates before and after discontinuation of first-line treatment might lead to over-estimation of second-line treatment costs. We investigate the importance of this potential bias through sensitivity analysis on the duration of second-line treatments.

4.3.4 Treatment effectiveness and extrapolation

To apply the partitioned survival model described above, OS, PFS and TTD curves are required for cabozantinib and comparators, extrapolated over the 20-year time horizon. The company's approach to estimating these curves is described in section B.3.3 of the CS. They present two sets of base case results:

1. **Direct comparison** (cabozantinib vs. sunitinib)
   This analysis is based on patient-level data from the CABOSUN trial, with OS, PFS and TTD curves separately fitted for cabozantinib and sunitinib arms using six families of survival functions: exponential, Weibull, Gompertz, lognormal, loglogistic and generalised gamma. For their direct base case, the company chose an exponential distribution for OS and lognormal distributions for PFS and TTD.
2. **Indirect comparison** (cabozantinib vs. sunitinib and pazopanib)
   ITC meta-analyses were conducted to fit PFS and OS curves to regenerated KM

data from CABOSUN and COMPARZ, as discussed in section 3.1.7 above. Two methods were used:

a. ITC parametric curves, fixed and random effect models for five survival functions: exponential, Weibull, Gompertz, lognormal, loglogistic (Ouwens et al. method).[36] The generalised gamma distribution was not implemented due to the lack of the incomplete gamma function in WinBUGS software. The company reports that treatment was tested as a covariate, but the model only includes curves that were fitted separately for cabozantinib and sunitinib.
b. ITC fractional polynomial (FP) curves, fixed effect, with five first-order and five second order functions (Jansen method).[40] For their ITC base case, the company chose the second-order FP model with P1=P2=-1 for PFS and OS.

As TTD KM plots are not available for COMPARZ, the company uses the CABOSUN lognormal curves for cabozantinib and sunitinib, and assumes the latter would also apply to pazopanib.

We describe and critique the company's choice of OS, PFS and TTD curves below. Further critique and explanation for the ERG's preferred approach is given in section 4.4.1.

4.3.4.1 Overall Survival (OS)

OS direct comparison

The company's preferred model for OS is the exponential, with Weibull and Gompertz tested in scenario analysis. They state that this decision was based on the Survival Model Selection Process (SMEEP) from NICE Decision Support Unit Technical Support Document 14.[46]

- **Proportional hazards (PH):** The company states that PH does not hold for OS in CABOSUN. This is apparent from the KM plots (CS B.2.6 Figures 6 and 7) which cross, and the Schoenfeld and log cumulative hazard plots support this conclusion (response to clarification question A3).
- **Goodness-of-fit (AIC/AICC/BIC):** Statistical measures of fit for OS are shown in CS Tables 34 and 35. There is inconsistency between treatments. For cabozantinib, the lognormal has the best BIC, followed closely by exponential and loglogistic. But for sunitinib, Gompertz has the best BIC followed by exponential and Weibull. The company uses the exponential for both arms in their base case, arguing that this has a reasonable fit for both cabozantinib and sunitinib.

- **Plausibility of extrapolation:** The company states that visual inspection of the curves by clinical oncologists led to the conclusion that the lognormal, loglogistic and gamma distributions give unrealistically optimistic long-term survival.

We show the fitted curves together with CABOSUN KM data in Figure 15 and selected summary statistics in Table 13 below. The ERG agrees that the exponential has a reasonable visual and statistical fit for both treatments and that it yields plausible estimates of long-term survival: 13% at five years for sunitinib in comparison with 21% for an observational cohort from the IMDC dataset that includes patients with a better risk profile.[47] Use of an exponential distribution for both treatments conflicts with the conclusion that OS hazards are not proportional. But we suggest that the exact shape of the CABOSUN KM curves should not be over-interpreted given the modest sample size (n=157) and lack of explanation for why the curves should come together and then diverge between about 13 and 20 months. The Weibull distribution and Gompertz provide reasonable alternatives for scenario analysis.

The ERG is concerned that the OS curves appear to have been fitted to CABOSUN January 2017 data cut, rather than the most recent July 2017 dataset which was less favourable for cabozantinib (CS B.2.6 Figures 6 and 7). The CS does not state which dataset was used, but the January 2017 KM plot is reproduced in the economic chapter (CS B.3.3 Figure 13) and KM data provided by the company in response to a clarification question also relates to this earlier cut-off. Failure to use the most recent available data will introduce bias in favour of cabozantinib. We consider this issue in ERG additional analysis; section 4.4.1 below.

OS indirect comparisons

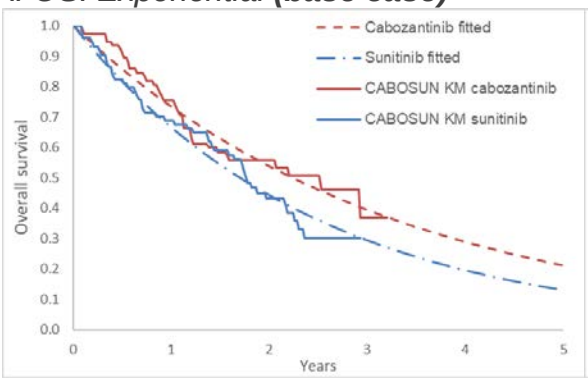
Figure 16 shows the ITC parametric and best-fitting FP survival curves in relation to the CABOSUN KM curves. We omit the COMPARZ KM curves from these graphs for clarity; but note that they are similar to the CABOSUN KM curve for cabozantinib and lie above the CABOSUN KM curve for sunitinib. This reflects the better risk status of participants in COMPARZ than in CABOSUN. The summary OS statistics are in Table 13 below.

The company use a second order FP model with P1=P2=-1 for OS in their ITC base case and three random effect parametric curves (exponential, Weibull and Gompertz) and two FPs (P1=-0.5, P2=0) and (P1=-1, P2=0) in scenario analysis. Their rationale for this choice is outlined in the CS:

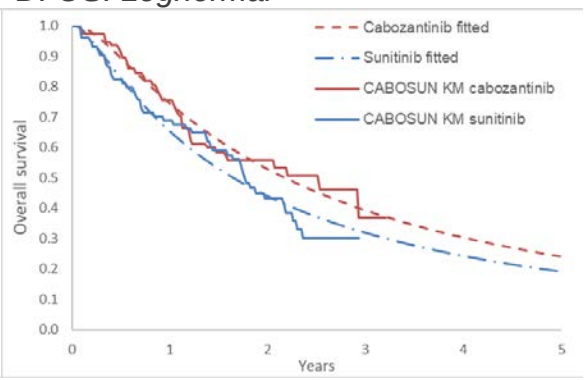
- **Proportional hazards (PH):** The company conclude that the proportional hazards assumption holds for OS in COMPARZ but not in CABOSUN. The ERG agrees with this conclusion.
- **Goodness-of-fit (DIC):** Measures of statistical fit are shown in CS Tables 23 and 24 (B.2.9). The company state that they selected the second-order FP with P1=P2=-1 because it had the best DIC statistic. They note that the first-order FPs have higher DIC statistics than second-order models (clarification question A25), so are not used in scenario analysis.
- **Plausibility of extrapolation:** The company state that two of the second-order FP models (P1=-1, P2=0.5) and (P1=-1, P2=1) are not recommended because they have "unreasonably flat tails". We note that this can also be said of the lognormal and loglogistic parametric models.

There is uncertainty over the robustness of the ITC results due to differences in the trial populations. The CABOSUN OS KM curves are also noisy, reflecting the small sample size and relative immaturity of the data. This makes it difficult to assess the fit and extrapolation of the 20 ITC curves included in the model. We consider that the RE exponential and FP P1=P2=-1 OS curves are both reasonable, with no clear reason to choose between them. The Weibull appears similar but with rather lower estimates of long-term survival with standard treatment. Conversely, the lognormal and loglogistic curves and two FP curves that the company includes in scenario analysis give high estimates of long-term survival, which we consider unrealistic. We therefore focus on the RE exponential, FP P1=P2=-1 and Weibull functions for OS in ERG additional analysis. We also consider the likely effect of using the most recent OS data from CABOSUN (July 2017 cut-off) to model cost-effectiveness.

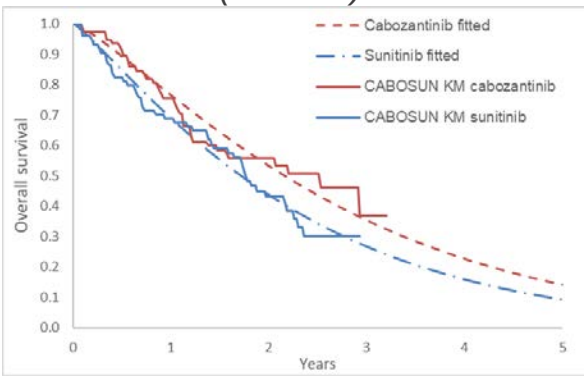
A: OS: Exponential ***(base case)***



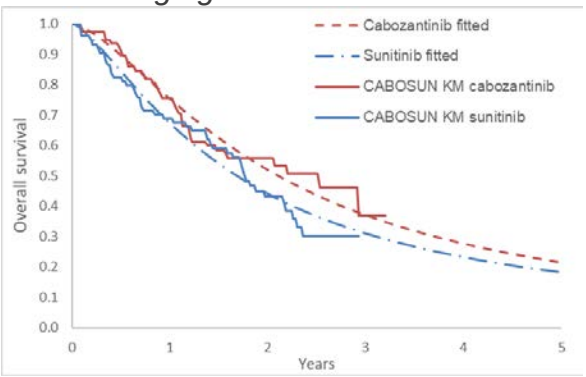
D: OS: Lognormal



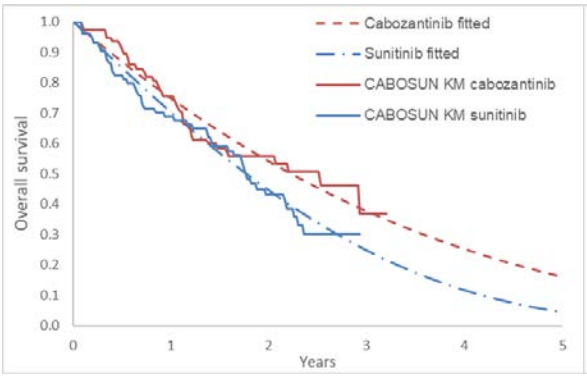
B: OS: Weibull (scenario)



E: OS: Loglogistic



C: OS: Gompertz (scenario)



F: OS: Generalised gamma

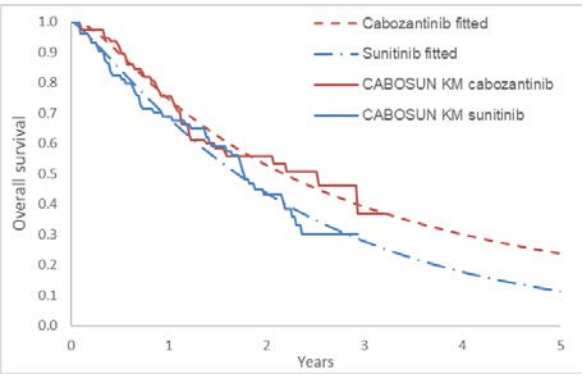
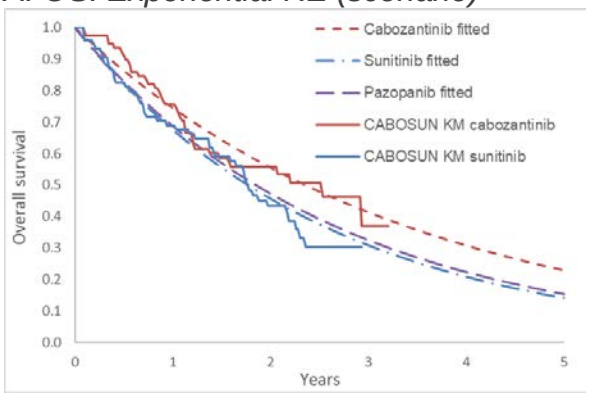


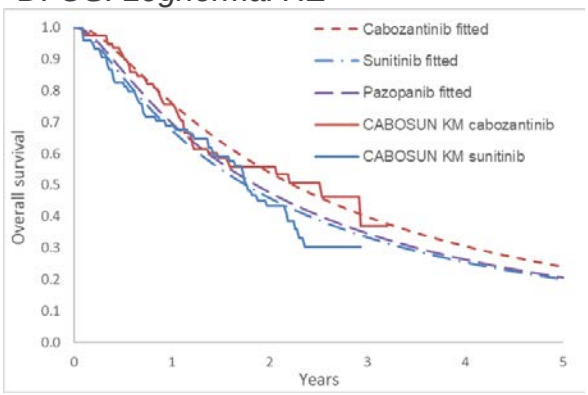
Figure 15 OS curves - fitted to CABOSUN data (direct comparison)

Source: Figures generated by ERG from company model and KM data used to fit models (Jan 2017 data cut-off)

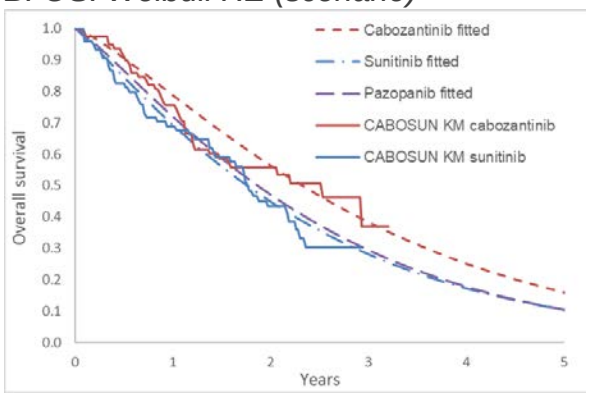
A: OS: Exponential RE (scenario)



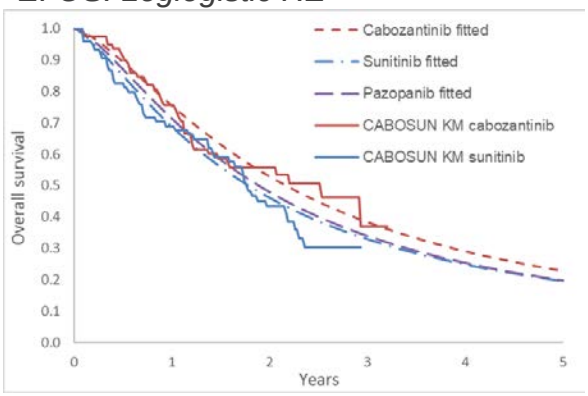
D: OS: Lognormal RE



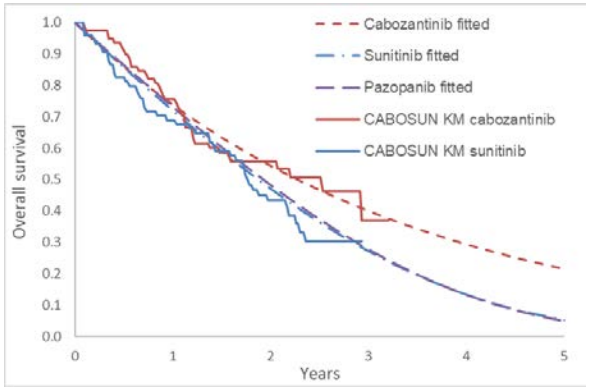
B: OS: Weibull RE (scenario)



E: OS: Loglogistic RE



C: OS: Gompertz RE (scenario)



F: OS: FP P1=P2=-1 ***(base case)***

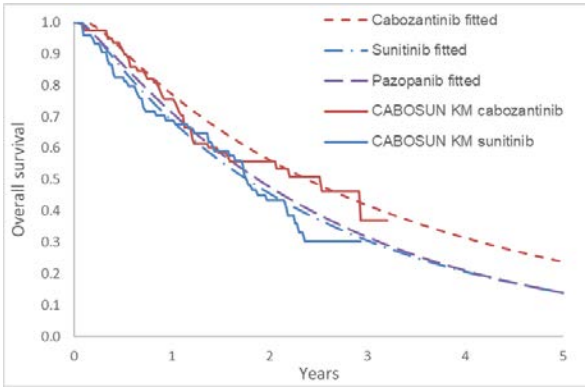


Figure 16 OS curves – ITC models fitted to CABOSUN AND COMPARZ

Source: Figures generated by ERG from company model and KM data used to fit models (Jan 2017 data cut-off). RE= Random effects

Table 13 Summary statistics for OS curves

	Model fit statistics [a]		Median OS (months)			5-year OS (%)		
			Cabo	Suni	Pazo	Cabo	Suni	Pazo
Data sources								
CABOSUN (Jan 17)	-		30.3	21.0	-			
CABOSUN (July 17)	-		26.6	21.2	-			
COMPARZ [b]	-		-	29.3	28.4			
IMDC [c]	-		-	22.3	22.6	-	21%	24%
Tivozanib STA [d]	-		-	27.5	29.2			
Fitted models: direct comparison (CABOSUN)								
Exponential	358.3	398.6	27.1	20.6	-	21%	13%	-
Weibull	360.7	401.9	26.1	20.8	-	14%	9%	-
Gompertz	362.5	394.5	26.9	21.4	-	16%	4%	-
Lognormal	358.3	403.0	26.3	20.0	-	24%	19%	-
Loglogistic	358.7	403.0	25.5	20.4	-	22%	18%	-
Gamma	362.6	406.2	26.3	20.5	-	24%	11%	-
Fitted models: ITC parametric random effects (CABOSUN & COMPARZ)								
RE Exponential	1768.9		28.6	21.4	22.4	23%	14%	15%
RE Weibull	1757.2		28.2	21.4	22.7	16%	11%	11%
RE Gompertz	1775.0		27.6	22.8	23.4	22%	6%	5%
RE Lognormal	1713.2		27.2	21.4	22.6	24%	20%	21%
RE Loglogistic	1733.4		26.3	21.7	22.9	23%	20%	20%
Fitted models: ITC fractional polynomials (CABOSUN & COMPARZ)								
FP P=-1	1722.8		27.8	21.6	22.6	18%	12%	12%
FP P=-0.5	1739.5		27.8	21.7	22.8	17%	11%	11%
FP P=0	1757.7		27.5	22.0	23.3	17%	10%	10%
FP P=0.5	1769.0		27.7	22.2	23.5	18%	8%	8%
FP P=1	1773.0		28.0	22.3	23.6	21%	7%	6%
FP P1=-0.5, P2=0	1716.5		28.4	20.8	22.9	23%	19%	19%
FP P1=-1, P2=0	1713.9		28.6	21.2	23.0	23%	17%	16%
FP P1=-1, P2=-1	1711.9		29.0	21.5	22.8	24%	14%	14%
FP P1=-1, P2=0.5	1716.2		29.0	21.3	23.1	28%	15%	15%
FP P1=-1, P2=1	1718.3		29.3	21.5	23.1	34%	14%	13%

a As reported in CS Tables 23, 24, 34 and 35: direct comparison Bayesian Information Criterion (BIC) for cabozantinib / sunitinib; parametric ITC: Deviance Information Criterion (DIC) for fixed/random effects models; and ITC FPs: Deviance Information Criterion (DIC) for first/second order models

b Motzer et al. 2013 analysis of COMPARZ trial data.[34]

c Ruiz-Morales et al. analysis of 7438 patients with metastatic RCC treated at first line with sunitinib (n=6519) or pazopanib (n=919)[47]

d ERG preferred results from Tivozanib STA (TA512)[22]

4.3.4.2 Progression free survival (PFS)

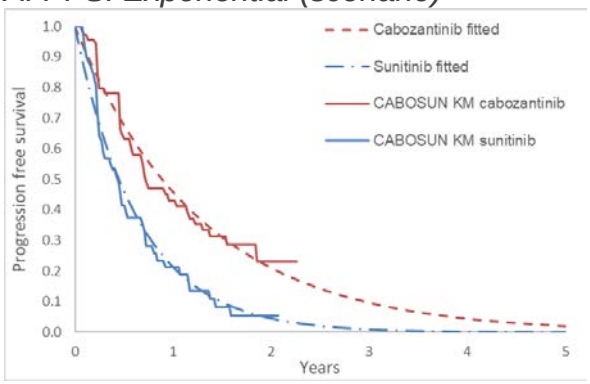
PFS direct comparison

The KM plot of PFS from CABOSUN is shown in CS B.3.3 Figure 14. We show selected graphs comparing the company's fitted curves with the CABOSUN KM plots in Figure 17 and summary statistics in Table 14 below. The KM plots from COMPARZ are higher than the KM for the sunitinib arm in CABOSUN. This is expected given the lower risk status of the COMPARZ population.

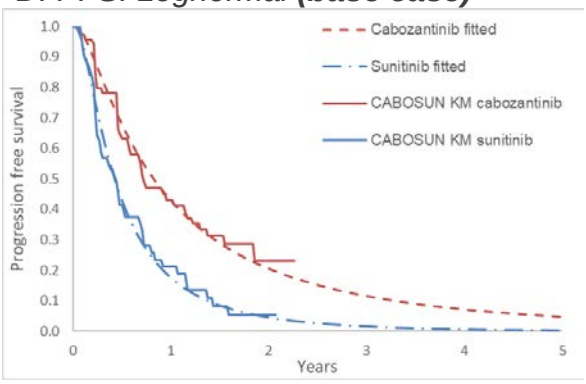
The company use separately fitted lognormal distributions for PFS in their direct base case analysis, with exponential, Weibull and Gompertz distributions in scenario analysis. They state that they made this choice based on the following considerations:

- **Proportional hazards (PH):** The company argues that the PH assumption is not appropriate for PFS in CABOSUN (CS Appendix D, Table 12). However, the ERG considers that this conclusion was not supported by the proportionality test or by the Schoenfeld and log-cumulative hazard plots (see section 3.1.7.3 above).
- **Goodness-of-fit (AIC/AICC/BIC):** Evidence of the fit of the parametric curves to trial data is provided with AIC, AICC and BIC statistics in Tables 36 and 37 of the CS (B.3.3). These show that for both study arms, the lognormal distribution provides the best fit to PFS data, followed by generalised gamma and loglogistic distributions.
- **Plausibility of extrapolation:** The company states that plausibility was assessed by visual inspection of the curves by oncologists currently practising in the NHS and England based on their clinical experience. No further information is provided about how this clinical assessment of validity was conducted or how it informed the choice of PFS curve.

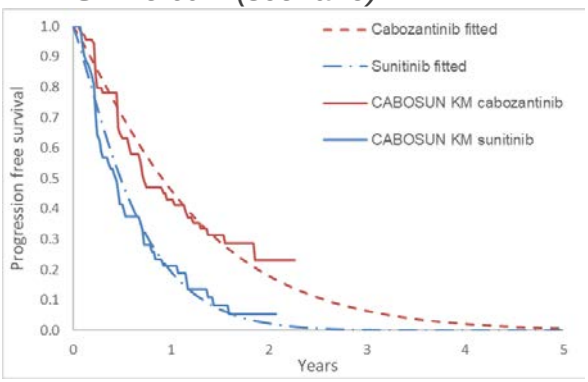
A: PFS: Exponential (scenario)



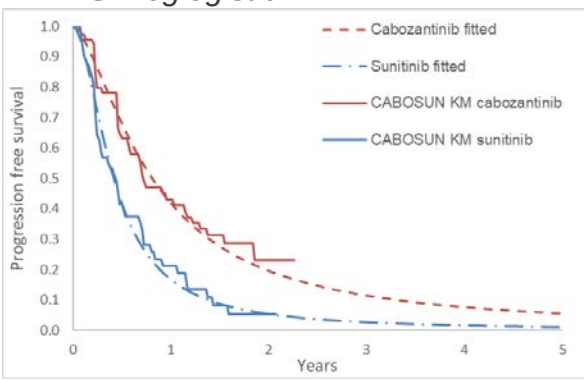
*D: PFS: Lognormal **(base case)***



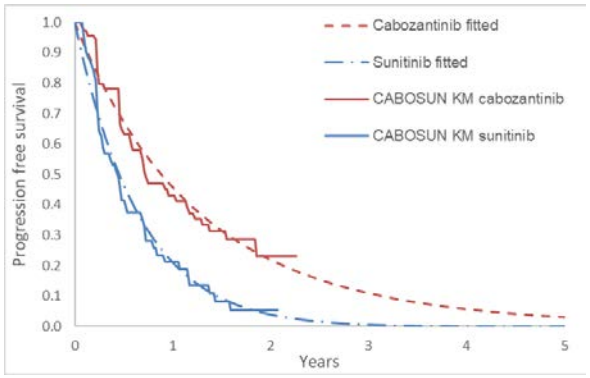
B: PFS: Weibull (scenario)



E: PFS: Loglogistic



C: PFS: Gompertz (scenario)



F: PFS: Generalised gamma

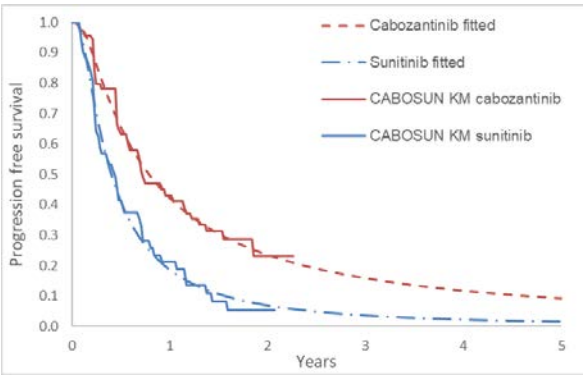


Figure 17 PFS curves - fitted to CABOSUN data (direct comparison)

Source: Figures generated by ERG from company model and KM data

We agree that the lognormal, exponential and Gompertz functions show a reasonable visual fit to the trial data, although they both overestimate median PFS for cabozantinib (as do all the other functions due to a 'dip' in the PFS KM curve). The Weibull has a poor visual fit.

PFS indirect comparison

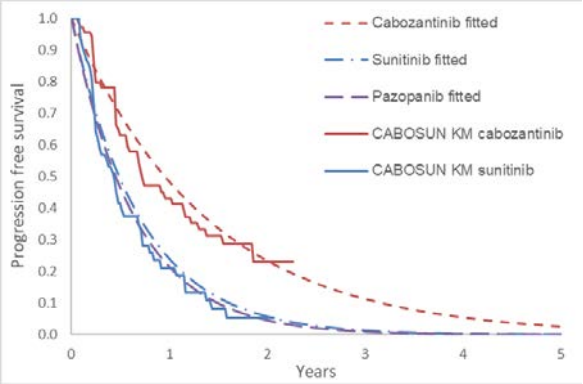
ITC curves for PFS are shown in Figure 18 alongside the CABOSUN KM plots.

The company use the second-order FP model with P1=P2=-1 for their ITC base case, and exponential, Weibull and Gompertz random effect models in scenario analyses. This choice was based on the following considerations:

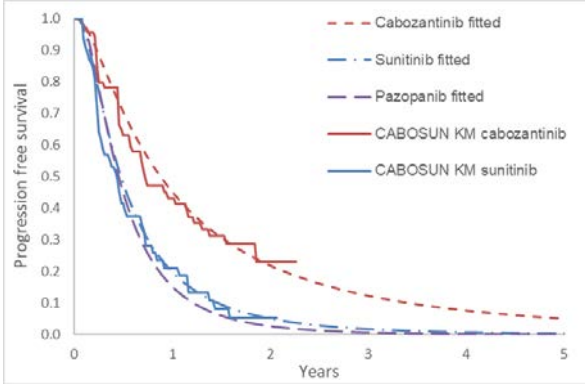
- **Proportional hazards (PH):** The company concludes that the PH assumption was violated for PFS in both CABOSUN and COMPARZ (Appendix D, Table 12). As noted above, we question this conclusion for CABOSUN. But for COMPARZ, the Schoenfeld and log cumulative hazard plots do suggest a change in hazard ratio over time (section 3.1.7.3 above).
- **Goodness-of-fit (DIC):** Evidence of the fit of the parametric and FP curves to the CABOSUN and COMPARZ data is provided with DIC statistics in Tables 23 and 24 of the CS (section B.2.9). For the parametric models, the lognormal distribution had the lowest DIC with both, followed by loglogistic. Among the FP models, the second-order P1=P2=-1 model had the lowest DIC statistic.
- **Plausibility of extrapolation:** The company notes that they decided not to include other second-order FP models in scenario analyses because they predict a high PFS rate at year 5. There is no discussion of the plausibility of the other parametric models, or of the long-term continuation of treatment effects observed in the trials.

Summary statistics in Table 14 show that all the ITC models overestimate median PFS for cabozantinib in relation to the CABOSUN result and several also overestimate median PFS with sunitinib. Long-term projections of PFS also seem optimistic for some models, particularly the second-order FP models and the Gompertz parametric model. Although the exponential and Weibull models do not have this problem, they have a worse visual fit and large overestimates of median PFS. We conclude that the lognormal and loglogistic models seem to provide the best balance of fit to the CABOSUN data with a realistic long-term extrapolation.

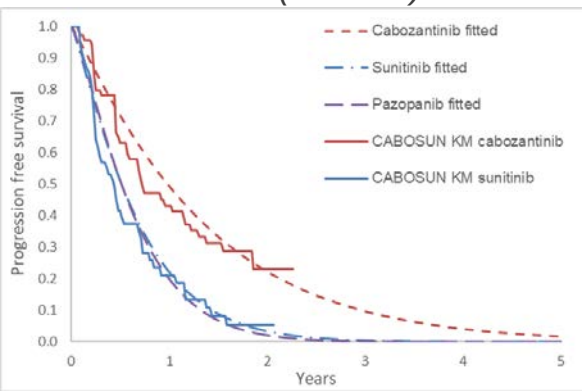
A: PFS: Exponential RE (scenario)



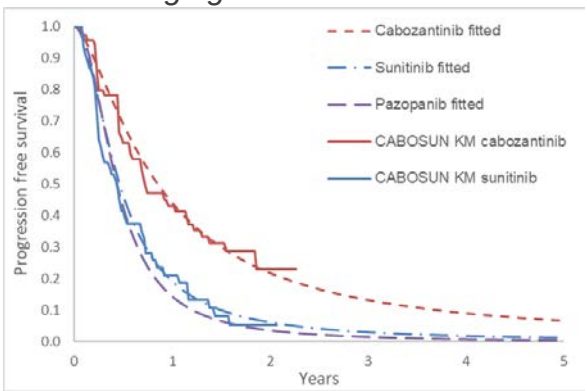
D: PFS: Lognormal RE



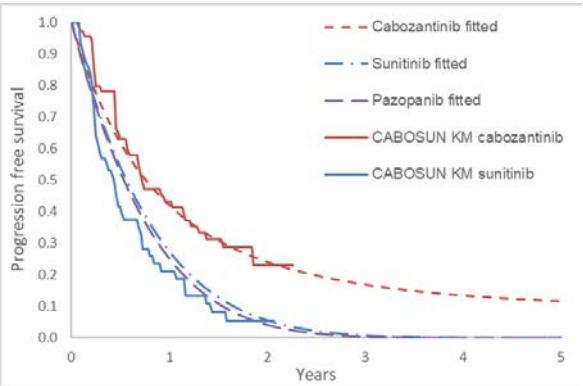
B: PFS: Weibull RE (scenario)



E: PFS: Loglogistic RE



C: PFS: Gompertz RE (scenario)



F: PFS: FP P1=P2=-1 ***(base case)***

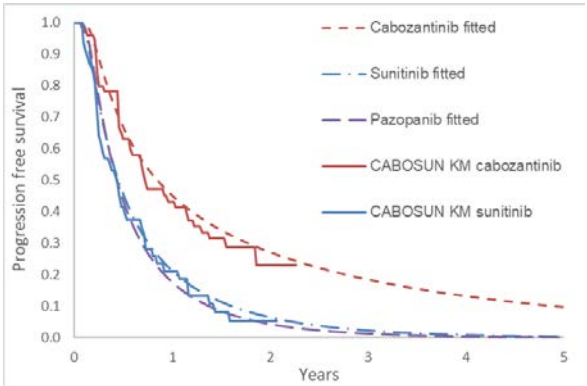


Figure 18 PFS curves – ITC models fitted to CABOSUN AND COMPARZ

Source: Figures generated by ERG from company model and KM data

Table 14 Summary statistics for PFS curves

	Model fit statistics [a]		Median PFS (months)			5-year PFS (%)		
			Cabo	Suni	Pazo	Cabo	Suni	Pazo
Data sources								
CABOSUN	-		8.6	5.3	-			
COMPARZ [b]	-		-	9.5	8.4			
IMDC data [c]	-		-	8.4	8.3	-	5%	8%
Tivozanib STA[d]	-		-	6.8	8.4			
Fitted models: direct comparison (CABOSUN)								
Exponential	325.6	303.2	10.9	5.6	-	2%	0%	-
Weibull	328.5	304.8	11.2	6.1	-	1%	0%	-
Gompertz	329.8	307.3	10.9	5.7	-	3%	0%	-
Lognormal	321.7	295.8	10.2	5.4	-	5%	0%	-
Loglogistic	323.6	298.0	10.0	5.3	-	5%	1%	-
Gamma	324.6	298.8	9.7	5.1	-	9%	1%	-
Fitted models: ITC parametric random effects (CABOSUN & COMPARZ)								
RE Exponential	1941.6		11.7	6.1	5.6	3%	0%	0%
RE Weibull	1945.5		12.0	6.4	6.3	2%	0%	0%
RE Gompertz	1943.5		9.3	7.0	6.6	11%	0%	0%
RE Lognormal	1860.6		10.8	5.9	5.6	5%	0%	0%
RE Loglogistic	1887.8		10.4	5.7	5.4	7%	1%	1%
Fitted models: ITC fractional polynomials (CABOSUN & COMPARZ)								
FP P=-1	1910.4		11.9	6.6	6.4	1%	0%	0%
FP P=-0.5	1932.0		12.0	6.7	6.5	1%	0%	0%
FP P=0	1945.9		11.7	6.6	6.5	2%	0%	0%
FP P=0.5	1947.6		11.4	6.4	6.3	3%	0%	0%
FP P=1	1943.6		11.2	6.2	5.9	6%	0%	0%
FP P1=-0.5, P2=0	1852.1		10.5	5.4	5.3	14%	4%	3%
FP P1=-1, P2=0	1840.3		10.5	5.3	5.2	12%	3%	2%
FP P1=-1, P2=-1	1825.0		10.4	5.6	5.4	10%	0%	0%
FP P1=-1, P2=0.5	1850.4		10.3	5.5	5.3	17%	4%	3%
FP P1=-1, P2=1	1858.1		10.2	5.7	5.5	21%	5%	3%

a As reported in CS Tables 23, 24, 34 and 35: direct comparison Bayesian Information Criterion (BIC) for cabozantinib / sunitinib; parametric ITC: Deviance Information Criterion (DIC) for fixed/random effects models; and ITC FPs: Deviance Information Criterion (DIC) for first/second order models

b Motzer et al. 2013 analysis of COMPARZ trial data.[34]

c Ruiz-Morales et al. analysis of 7438 patients with metastatic RCC treated at first line with sunitinib (n=6519) or pazopanib (n=919).[47]

d ERG preferred results from Tivozanib STA (TA512)[22]

4.3.4.3 Time to discontinuation (TTD)

Time on treatment is modelled for cabozantinib and sunitinib based on parametric curves fitted to CABOSUN data (CS B.3.3 Figure 15). For pazopanib, no TTD data were available, so the company assume that the sunitinib curve would also apply to pazopanib. They justify this by noting that the mean treatment duration in COMPARZ was 11.5 months for both treatments. The median duration of treatment was also similar: 7.6 months (range 0 to 38) for sunitinib and 8.0 months for pazopanib (range 0 to 40) respectively.[35] The ERG agrees with this approach.

We illustrate the fitted TTD curves alongside KM plots (digitised by the ERG) in Figure 19 below. The visual fit appears similar for the different parametric functions and as the TTD data are mature, extrapolation is less of an issue than for PFS and OS. The model fit statistics are shown in Table 38 and 39 of the CS (B.3.3). The optimum curve differs by measure of fit and by treatment. The company use the lognormal in their base case analysis, as they argue that this provides a good fit for both cabozantinib and sunitinib. They also test exponential, Weibull, Gompertz and generalised gamma in scenario analysis. There is no obvious reason for excluding the loglogistic from scenario analysis.

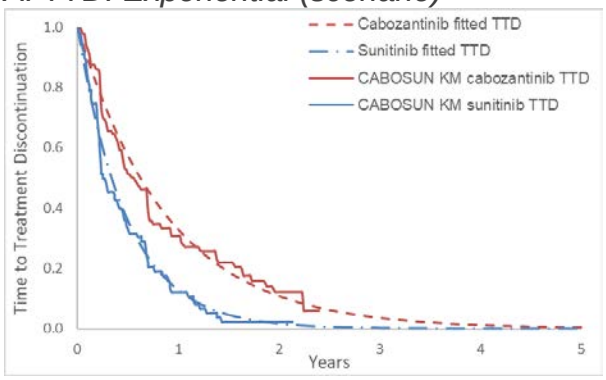
Table 15 below summarises the survival curves used for OS, PFS and TTD.

Table 15 Survival curves used in company analyses

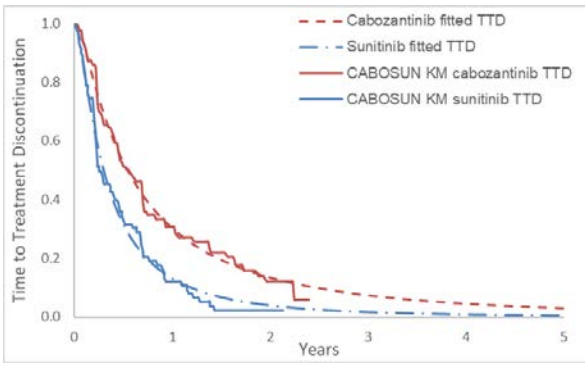
Curve	Method	Treatment	CS Base case	CS scenarios
OS	**Direct** CABOSUN	Cabozantinib Sunitinib	Exponential	Weibull Gompertz
	ITC CABOSUN & COMPARZ	Cabozantinib Sunitinib Pazopanib	FP P1=P2=-1	RE exponential RE Weibull RE Gompertz FP P1=-0.5, P2=0 FP P1=-1, P2=0
PFS	**Direct** CABOSUN	Cabozantinib Sunitinib	Lognormal	Exponential Weibull Gompertz
	ITC CABOSUN & COMPARZ	Cabozantinib Sunitinib Pazopanib	FP P1=P2=-1	RE exponential RE Weibull RE Gompertz
TTD	**Direct** CABOSUN	Cabozantinib Sunitinib Pazopanib	Lognormal	Exponential Weibull Gompertz Generalised gamma

RE = Random effects; FP = Fractional polynomial

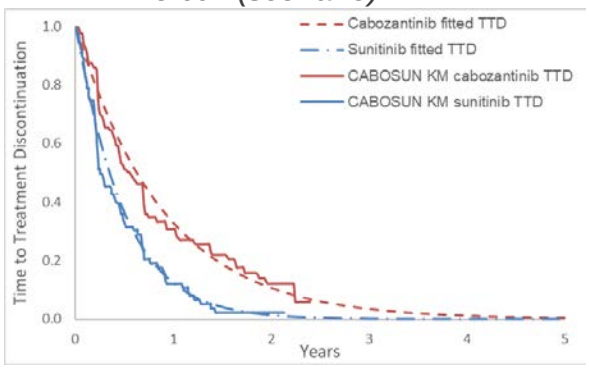
A: TTD: Exponential (scenario)



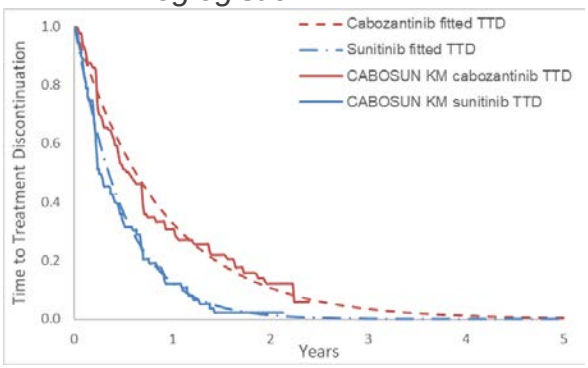
D: TTD: Lognormal ***(base case)***



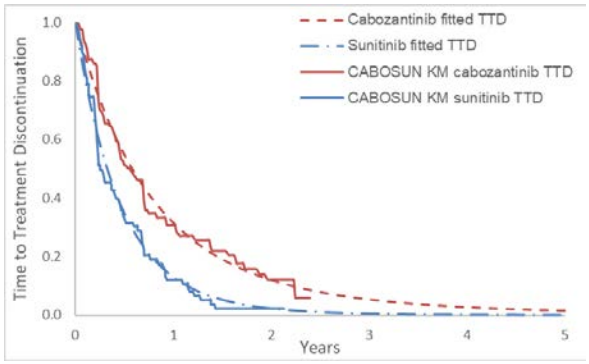
B: TTD: Weibull (scenario)



E: TTD: Loglogistic



C: TTD: Gompertz (scenario)



F: TTD: Generalised gamma (scenario)

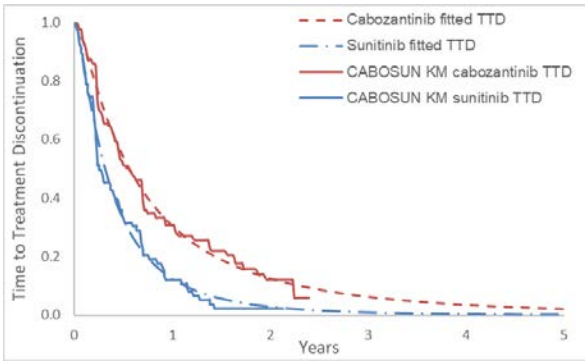


Figure 19 TTD curves - fitted to CABOSUN data (direct comparison)

Source: Figures generated by ERG from company model and digitised KM from CS B.3.3 Figure 15

4.3.5 Health related quality of life

For calculation of QALYs, the model requires estimates of utilities for the two health states (PF and PD) and disutilities for grade 3 or 4 treatment emergent adverse events (TEAEs).

EQ-5D or other relevant utility data was not collected in CABOSUN. The CS describes a post hoc analysis using a quality-adjusted time without symptoms of disease or toxicity (Q-TWiST) framework. Outputs from this analysis are not fed into the economic model. As they fall outside the NICE reference case, we do not discuss them further.

The company conducted a systematic search to identify utility values that could be used in the model. The company's search strategy is described in CS Appendix H. The ERG considers this search strategy to be adequate and up to date (see section 3.1.1). The search identified 22 publications which reported EQ-5D-based utilities relevant to first-line treatment of advanced or metastatic RCC. Of these publications, the company deemed four to be relevant to a UK setting; summarised in CS Table 40. Of these four publications, only Swinburn et al. 2010[15] reported health state-specific utilities that would be suitable for inclusion in the model. EQ-5D utility values reported in Swinburn et al. 2010 include; stable disease with no adverse event (0.795) and disease progression (0.355). In addition, the company checked and reported relevant utility values from previous NICE submissions. The utility values used in the company base case and scenarios are summarised in Table 16 below.

Table 16 Utility values (adapted from CS Tables 46 & 47)

Health state	Utility value: mean (SE)	95% CI	CS reference
Progression free			
Base case	0.726 (0.011)	0.705 to 0.748	Tivozanib TA512 [22]
Scenario	0.70 (0.01)	0.680; 0.720	Pazopanib TA215[48]
Scenario	0.78 (0.078)*	0.627; 0.933*	Sunitinib TA169[49]
Scenario	0.795 (0.0176)	0.761; 0.830	Swinburn 2010[15]
Progressed disease			
Base case	0.649 (0.019)	0.612 to 0.686	Tivozanib TA512 [22]
Scenario	0.59 (0.059)*	0.474; 0.706*	Pazopanib TA215[48]
Scenario	0.705 (0.071)*	0.567; 0.843*	Sunitinib TA169[49]
Scenario	0.355 (0.0288)	0.299; 0.412	Swinburn 2010[15]
TEAE grade 3/4			
Base case	-0.2044 (0.0682)	-0.0707 to -0.3381	COMPARZ[50]
Scenario	-0.0550 (0.0068)	-0.0418; -0.0685	METEOR trial[51]

Abbreviations: TEAE, treatment emergent adverse effects; CI, confidence interval; SE, standard error; *SE or 95% CI not available in literature; 10% of the mean assumed.

4.3.5.1 Health state utilities

The ERG agrees that the health state utility values in Table 16 meet the NICE reference case and are suitable for inclusion in the model. The values for the progression free health state are reasonably consistent (0.70 to 0.80). However, there are large differences between the available utility estimates after disease progression (0.36 to 0.71). The Swinburn et al.[15] study gives the biggest difference in utility between the PF and PD state, a loss of 0.44. This study used time trade-off approach to elicit UK societal preferences for health states associated with metastatic RCC, rather than EQ-5D valuations. The ERG agrees with the company's preference for the health state utility values used in the tivozanib STA, with scenarios based on the pazopanib and sunitinib STA utility values and Swinburn et al.[15]

We spotted a disparity between the utility scores from the sunitinib STA as reported in CS Tables 43 and 47 and the values applied in the company's model. This error did not affect the results for the company's base case, but the results for the scenario analysis with sunitinib health state utilities in CS Tables 66 and 67 are incorrect. The company corrected these errors in response to a clarification question (B4).

4.3.5.2 Adverse event disutilities

As adverse event specific utility data was not reported in CABOSUN, the company derived base case TEAE disutilities from the COMPARZ study, with values from the METEOR trial[51] in scenario analyses. The company assumes that TEAE disutilities are not disease-specific and that all types of grade 3 or 4 events elicit the same utility loss for a fixed period of 4 weeks and a fixed number of episodes per patient per TEAE. These assumptions yield a mean QALY loss of 0.0225 per TEAE in the base case (0.006 in the METEOR trial-based scenario). The ERG considers that it is highly unlikely that the QALY loss is the same for all types of TEAE, but that these assumptions reflect a reasonable average. We conduct additional scenario analyses to test model sensitivity to the TEAE disutility parameter, including higher as well as lower estimates of the disutility.

The company models the incidence of grade 3/4 TEAEs based on reported rates from the CABOSUN trial for cabozantinib and sunitinib and from COMPARZ for pazopanib, see Table 17 below. Only events with a reported incidence of 5% or greater in at least one arm were included. We note that of 59 types of adverse events listed in the company's model, only 18 events with incidences equal to or greater than 5% were modelled. We test the impact of changing the inclusion threshold for TEAEs in scenario analysis.

The model does not include QALY decrements for TEAEs associated with second line treatments. We consider that this is reasonable, as utility loss related to subsequent treatments should be reflected in the PD health state utility. This may also be true for the PF health state utility – thus there may be some degree of double counting due to the inclusion of TEAE disutilities for the first-line treatments. However, it is important to reflect potential disutility related to differences in adverse effect incidence for main treatments of interest.

Table 17 Incidence of modelled grade 3/4 adverse events by treatment and study

Adverse Event	Cabozantinib (n=78) CABOSUN (%)	Sunitinib (n=72) CABOSUN (%)	Pazopanib (n=554) COMPARZ (%)
Decreased appetite	5	1	1
Diarrhoea	10	11	9
Dyspnoea	1	6	3
Embolism	8	-	-
Fatigue	6	17	11
Hyperglycaemia	-	6	5
Hypertension	28	21	15
Hyponatremia	9	8	7
Hypophosphatemia	9	7	4
Hypotension	5	1	-
Increased ALT	5		17
Increased AST	3	3	12
Lymphocytopenia	1	6	5
Neutropenia	-	3	5
Pain	5	-	-
Hand and foot syndrome	8	4	6
Stomatitis	5	6	1
Thrombocytopenia	1	11	4

Source: Extracted from the model by the ERG.

4.3.6 Resource use and costs

The costs included in the economic model consist of drug acquisition and administration costs for first-line and subsequent treatments (adjusted for dose intensity and wastage where appropriate), health state management costs (for PF and PD), costs incurred for the management of adverse event costs and costs incurred at the end of life.

The company conducted a systematic literature review to identify published resource use and cost data relevant to the cost-effectiveness analysis. From a total of 61 full-text articles identified, the company judged 22 studies to be eligible for data extraction since they related to countries in the company's scope (European countries, Australia and Canada). CS Appendix I provides a detailed description of the company's search strategy and inclusion criteria. The ERG considers that the company's literature review, which was updated in September 2017, is likely to reflect available evidence.

4.3.6.1 First-line drug costs

Table 18 summarises the drug acquisition costs for first-line treatments included in the company's model.

Table 18 Drug cost per week for first line treatments (adapted from CS Table 48)

Drug	Relative dose intensity (SE)	PAS discount	Cost per week without discount	Cost per week with discount
Cabozantinib	94.3% (1.5) [a]	[redacted]	[redacted]	[redacted]
Sunitinib	87.4% (6.3) [a]	First 6-week cycle free [c]	£457	£457
Pazopanib	86.0% (8.6) [b]	12.5% [b]	£450	£394

a CABOSUN CSR, Table 37, sunitinib mean relative dose intensity.[27]
b NICE pazopanib appraisal TA215[13]
c NICE sunitinib appraisal TA169.[12] Coded in model rather than as a simple discount

The cost per pack for all drugs are derived from the British National Formulary. The company base case includes published patient access schemes (PAS) discounts for pazopanib and sunitinib: 12.5% reduction on the list price for pazopanib (TA215)[13] and the first 6-week cycle free for sunitinib (TA169).[12] The company also applies a pre-existing confidential PAS discount of [redacted] for cabozantinib in previously-treated advanced RCC (TA463[14]) to the cost of first-line cabozantinib in their base case analysis.

Relative dose intensity is factored into the cost calculations to reflect the percentage of days with interrupted treatment – for example, due to adverse effects. Cabozantinib is available in doses of 20, 40 and 60 mg, with all doses priced equally. However, while a reduction in dose

does not impact on costs, an interruption in treatment may do so. The company bases estimates of dose intensity for cabozantinib and sunitinib on CABOSUN data and for pazopanib, estimates in the NICE appraisal TA215 are used. For comparison, the ERG analysis for the recent NICE tivozanib appraisal (TA512) included an 86% relative dose intensity for both sunitinib and pazopanib, based on values cited in previous NICE appraisals for these drugs. The appraisal committee in TA512 concluded that there is uncertainty over the impact of dose intensity on the cost of oral treatments, and that this is likely to fall somewhere between the ERG's estimates and 100%. We consider that the company's approach to modelling the cost impact of dose intensity is reasonable. However, we conduct an additional scenario analysis to test the effect of assuming the same relative dose intensities (86% and 100%) for all treatments.

The company did not include additional administration costs for oral chemotherapies in their model. The ERG agrees with this approach. We note that the NHS does incur costs for delivery of oral chemotherapies, which are included the National Schedule of Reference Costs (currency code SB11Z). However, the modelled health state costs (listed below) include a monthly consultant-led medical oncology outpatient visit and blood tests, which we assume would include the cost of procurement, prescribing and monitoring of oral chemotherapies.

4.3.6.2 Health-state costs

The CS reports assumptions about resource use and unit costs for disease management in Tables 49 and 50 (summarised in Table 19 below). The company assumes that patients have a monthly medical oncologist visit and blood tests and a computerised tomography (CT) scan every three months. For scenario analysis, they assume less frequent oncologist follow-up but with access to a specialist nurse. The company's model makes provision for second-line treatment following treatment failure (see below), with the same follow up and monitoring pre and post-progression.

The ERG considers that the company's estimates of health state costs are reasonable. They reflect resource use assumptions in previous NICE appraisals[12 13 22] and experts consulted by the ERG did not object to the company's assumptions, except that it was noted that in routine NHS care, patients would have some follow-up with a nurse specialist. Unit costs are based on appropriate and up-to-date national sources.[52 53]

Table 19 Health state management costs (adapted from CS Table 49 and 50)

Health state	Resource	Frequency per week		Unit cost	Reference
		Base case	Scenario		
PF	Outpatient (first)	Not applicable		£219	NHS Reference Costs 2016/17. Currency code WF01B, Service code 370, Medical oncology
	Outpatient (follow up)	0.25	0.08	£173	NHS Reference Costs 2016/20. Currency code: WF01A, Service code 370, Medical oncology
	Nurse visit	0	0.25	£173	Cost per hour. Nurse (GP practice), PSSRU Unit costs of health and social care 2016
	CT scan	0.08	0.08	£115	NHS Reference Costs 2016/17. Currency code: RD25Z. CT of three areas, without contrast
	Blood test	0.25	0.25	£3	NHS Reference Costs 2016/17. Currency code: DAPS05
PD	Outpatient (follow up)	0.25	0.08	£173	NHS Reference Costs 2016/17. Currency code: WF01A, Service code 370, Medical oncology
	Nurse visit	0	0.25	£173	Cost per hour. Nurse (GP practice), PSSRU Unit costs of health and social care 2016
	CT scan	0.08	0.08	£115	NHS Reference Costs 2016/17 Currency code: RD25Z Computerised Tomography Scan of three areas, without contrast
	Blood test	0.25	0.25	£3	NHS Reference Costs 2016/17. Currency code: DAPS05

Abbreviations: PF, progression free; PD, progressed disease; CT, computerised tomography.

4.3.6.3 Adverse event costs

The model includes costs for managing grade 3/4 TEAE with an incidence of ≥5% in the CABOSUN (cabozantinib and sunitinib) and COMPARZ (pazopanib) trials (see Table 17 above). The company's base case estimates of the costs of managing these events are summarised in Table 20. Resource use assumptions were derived from published estimates (CS Appendix I), HTA reports and clinical opinion. To avoid double counting, the company model omits costs for adverse events such as hyponatremia and hypotension which would be managed as part of regular follow up (included in the health state costs listed in the previous section). Unit costs come from standard national sources: including NHS Reference costs 2016/17, the British National Formulary (8/12/17) and the PSSRU Unit costs of Health and Social Care report 2016 (see Table 21).

The ERG finds the resources included in the CS to be comprehensive. We spotted certain textual errors in the CS: the unit cost for vascular ultrasound scan is wrongly reported as £75 and the cost of hospitalisation for lymphocytopenia is reported as £429. However, these errors do not affect cost-effectiveness results since the correct values are used in the model.

Table 20 Costs for management of adverse events (Adapted from CS Table 53)

Adverse event	Cost per event (£)	Assumptions about resource use
Diarrhoea	£567	Based on pazopanib NICE STA. Short stay admission and Loperamide 2 mg102 q.i.d 30 days
Dyspnoea	£68	Based on assumption of one pulmonologist visit
Embolism	£1,640	Based on NICE guidance on venous thromboembolic diseases: 1 ultrasound of coronary vessels. Therapy initiation with low molecular weight heparin for 6 months: deltaparin 18000 units o.d. units for first 30 days and continue with deltaparin 15000 units o.d. for further 5 months
Fatigue	£35	Based on tivozanib NICE STA. 20% of patients will have additional outpatient attendance
Hyperglycaemia	£156	Based on assumption: 1 visit to endocrinologists. Initiation of therapy with p.o anti-diabetic medication: metformin 500mg3 o.d. for one year
Hypertension	£128	Based on tivozanib NICE STA. 3 GP attendances, ramipril 5 mg + bendroflumethiazide 2.5 mg o.d. for 1 year
Lymphocytopenia	£362	Based on assumption of 20% of short stay emergency tariff (weighted average of SA35A-SA35E: £ 515) and 80% of day case tariff (weighted average of SA35B-SA35E: £ 288)
Neutropenia	£1,107	Based on the assumption: Granulocyte colony-stimulating factors (granulocyte CSF): Filgrastim. 5µg/kg for 14 days (dose is 450 µg o.d. for TM=90kg) Neupogen 30million units/1ml (1µg=100000 units)
Pain	£138	Based on assumption: Outpatient visit for pain management (CS Table 53 incorrectly cites monthly visits, but only one is costed in model).
Hand and foot syndrome	£104	Based on tivozanib NICE STA: 60% of patients will have additional outpatient attendance
Stomatitis	£42	Based on assumption: Local therapy for pain relief, local anaesthetics or other anti-inflammatory preparations - oral solution of dexamethasone 2mg/5ml
Thrombocytopenia	£351	Based on assumption: 20% of short stay emergency tariff (weighted average of SA12G-SA12K) and 80% of patients with day case tariff (weighted average of SA12G-SA12K)
Hyponatremia	£0	

Adverse event	Cost per event (£)	Assumptions about resource use
Hypophosphatemia	£0	Regular blood tests already considered under disease Management costs
Increased ALT	£0	
Increased AST	£0	
Hypotension	£0	Monthly outpatients visit already covered by disease management costs
Decreased appetite	£0	No stated justification in CS but not associated cost in company's model

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; od, once daily; qid, four times a day; STA, single technology appraisal

Table 21 Unit costs for management of adverse events (CS Table 51)

Health resource	Cost, £	Reference
Short stay admission due to diarrhoea	£558	HRG FD02E, Inflammatory Bowel Disease without Interventions, with CC Score 5+, NHS reference costs 2016/2017 [53]
Vascular ultrasound scan	£65 *	HRG RD47Z, NHS reference costs 2016/2017
Outpatient attendance for hand and foot syndrome	£173	HRG WF01A: service code 370 Medical oncology, NHS reference costs 2016/2017
Visit to endocrinologist due to hyperglycaemia	£146	WF01A, Service code 302, Endocrinology, NHS reference costs 2016/2017
GP visit due to hypertension	£36	GP visit-Unit cost per surgery consultation; PSSRU Cost of health and social care 2016[53]
Hospitalisation cost due to lymphocytopenia - Short stay emergency tariff	£492 *	HRG SA35A-SA35E short stay emergency tariff (weighted average), NHS reference costs 2016/2017
Hospitalisation cost due to lymphocytopenia - Day case	£330	HRG SA35A-SA35E day case tariff (weighted average), NHS reference costs 206/2017
Hospitalisation cost due to thrombocytopenia - Short stay emergency	£522	HRG SA12G-SA12K short stay emergency tariff (weighted average), NHS reference costs 2016/2017
Hospitalisation cost due to Thrombocytopenia - Day case	£308	HRG SA12G-SA12K day case tariff (weighted average), NHS reference costs 2016/2017
Outpatient attendance due to dyspnoea	£68	WF01A, Service code 342, Programmed Pulmonary Rehabilitation, NHS reference costs 2016/2017
Outpatient visit for pain management	£138	WF01A, Service code 191, Pain management, NHS reference costs 2016/2017

* Values from CS Table 51 corrected by ERG. Correct values were used in the submitted company model.

4.3.6.4 Second line treatment use and costs

The company's assumptions about the proportions of patients receiving subsequent treatments after failure of initial therapy are reported in Table 22. The base case reflects second-line treatments in the trials that inform model survival parameters: CABOSUN for cabozantinib and sunitinib; and COMPARZ for pazopanib. The company notes that although some of these treatments are not available or not approved for second line use in England, costing this mix of subsequent treatments is consistent with the implicit inclusion of their benefits through the trial estimates of survival. The company also conduct a scenario analysis in which the cost of subsequent treatment is adjusted to better reflect NHS practice. This scenario is largely based on ERG assumptions that were made for the NICE tivozanib appraisal.[22] For this current appraisal, the company adjust the distribution after first-line treatment with sunitinib or pazopanib, assuming that 10% of patients would start cabozantinib but only 40% axitinib.

Table 22 Distribution of subsequent treatments (Adapted from CS Table 56 and 57)

Second-line treatments	Following initial treatment with:					
	Company base case			Scenario analysis		
	Cabo[a]	Suni[a]	Pazo[b]	Cabo[c]	Suni[c]	Pazo[c]
Axitinib	23%	19%	6%	50%	40%	40%
Pazopanib	16%	12%	0%	0%	0%	0%
Sunitinib	13%	13%	29%	0%	0%	0%
Temsirolimus	9%	4%	6%	0%	0%	0%
Nivolumab	13%	15%	0%	30%	30%	30%
Everolimus	8%	19%	31%	10%	10%	10%
Sorafenib	1%	3%	11%	0%	0%	0%
Bevacizumab	0%	6%	7%	0%	0%	0%
Cabozantinib	1%	6%	0%	0%	10%	10%
Interferon	1%	0%	0%	0%	0%	0%
BSC	0%	0%	0%	10%	10%	10%

Abbreviations: cabo, cabozantinib; suni, sunitinib; pazo, pazopanib; BSC, best supportive care Sources:

a CABOSUN Clinical Study Report, Table 26.[27]
b COMPARZ Clinical Study Report [38]
c Tivozanib NICE STA[22] with 10% utilisation moved from axitinib to cabozantinib (following first-line treatment with sunitinib or pazopanib).

The ERG agrees with the company's general approach to modelling second-line treatments, with the base case following utilisation in the clinical trials and scenario analysis testing costs that are more reflective of NHS practice. This is not ideal, as the scenario omits the impact of NHS practice on survival, and the direction and magnitude of bias from this omission is unclear. However, we do not believe that it is feasible to model the effects of a different mix of second-line treatments. Re-analysis of OS data to adjust for second line treatment would not

be possible given the small sample size in CABOSUN: for example, only 4 patients in the cabozantinib arm and 6 in the sunitinib arm received nivolumab as subsequent therapy.[27] An alternative would be to explicitly model survival for the different second-line treatments after discontinuation of the initial therapy, but this would require a new model and systematic evidence review for the relevant population.

The mix of second-line treatments in the company's scenario analysis is similar to that used in the recent NICE appraisal for tivozanib. It excludes treatments that are not recommended by NICE for second-line use (sunitinib and sorafenib)[20] and those that have not been appraised for this indication (pazopanib, temsirolimus, bevacizumab and interferon-α). It includes three treatments recommended by NICE for second-line use: cabozantinib (TA463), nivolumab (TA417) and everolimus (TA432).[54] However, it does not include lenvatinib plus everolimus, which was also recently approved by NICE for second line use (TA498).[55] Clinical advice to the ERG is that current practice is usually to offer pazopanib or sunitinib at first-line, followed by nivolumab or cabozantinib at second line. If cabozantinib is made available at first-line, nivolumab or lenvatinib and everolimus would then probably be offered at second line. We conduct additional scenario analyses to test the impact of changing the costs of second line treatments.

Table 23 below (adapted from CS Tables 55 and 58) summarises the costs for second-line treatments included in the model. The company costs all second-line treatments at list price. This does not reflect current prices paid by the NHS, because agreed PAS discounts are in place for the five treatments approved by NICE for previously-treated advanced renal cell carcinoma. In ERG additional analyses below, we use the same approach and do not include PAS discounts for any second-line treatments. We apply all available PAS discounts in a confidential addendum to this report.

As for first-line treatments, the company assume that oral treatments at second-line do not incur an additional administration cost. We consider that this is reasonable, given that health management costs continue to include monthly outpatient visits after disease progression. The model includes administration costs for drugs delivered by injection (interferon-α) and IV infusion (nivolumab, bevacizumab and temsirolimus). The company assume that 25% of interferon-α injections are administered by a district nurse at a cost of £37 (£9.25 per dose). CS Table 59 states that the cost per IV infusion is £199, but the model actually applies a cost of £205 per infusion for bevacizumab and temsirolimus and omits the administration cost for nivolumab. We believe that this is an error and apply a cost of £205 per infusion for all drugs

delivered by IV infusion: this is the 2016/17 NHS Reference Cost for outpatient delivery of subsequent elements of the chemotherapy cycle (currency code SB15Z).

Table 23 Costs and duration of subsequent treatments (Adapted from CS Tables 55 and 58)

Subsequent treatments	Relative dose intensity, % (SE)	Cost per week, list price [a]	Duration (SE), weeks
Axitinib	102 (1.9)	£897	31.5 (3.2)
Pazopanib	86 (8.6)	£450	49.8 (5.0)
Sunitinib	87 (6.3)	£457	24.7 (2.5)
Temsirolimus	92 (9.2)	£825	17.0 (1.7)
Nivolumab	98 (9.8) plus 8% wastage	£1,572 *	42.0 (4.2)
Everolimus	84 (1.1)	£523	23.9 (2.4)
Sorafenib	80 (8.0)	£715	25.8 (2.6)
Cabozantinib	93.3 (9.3)	[redacted]	33.1 (3.3)
Bevacizumab	88 (8.8)	£1,050	24.0 (2.4)
Interferon-α	86 (8.6)	£155	12.0 (1.2)

[a] Cost at list price, including: adjustment for dose intensity; administration cost (£205 per infusion; £ and drug wastage (for drugs delivered by IV infusion or injection).

* Excludes cost of administration. This is corrected in ERG analysis.

The costs in Table 23 are also adjusted for relative dose intensity to account for missed doses of medications and wastage for vial formulations (nivolumab, bevacizumab and temsirolimus). The company also conducts a scenario analysis without wastage. We consider that wastage is likely to occur in clinical practice with vial formulations, and so should be included in the analysis.

The duration of subsequent treatments is based on a variety of sources, including NICE STAs for axitinib (TA333)[56] and nivolumab (TA417),[57] CABOSUN and COMPARZ for sunitinib and pazopanib, METEOR for cabozantinib and other trials for temsirolimus, everolimus, bevacizumab.[58 59 60]

4.3.6.5 End of life costs

The company base case includes a cost for end of life care applied in the last cycle before death. This comes from a 2014 Nuffield Trust report that estimated the cost of hospital care in the last three months of life for people within 2 years of a cancer diagnosis at £5,890. The company uprated this to £6208 at 2017 prices, based on general inflation indices.[61 62] The ERG believes that this is an under-estimate, due to the omission of costs for local-authority funded social care, district nursing and GP visits and the company's method of adjusting for inflation. Based on the Nuffield report, we estimate an end of life cost of £7,961 from an NHS

and PSS perspective and inflating using the Hospital and Community Health Services price index.[52] We include this revised figure in ERG analyses, but also conduct a scenario analysis excluding end of life care costs.

4.3.7 Model validation

The company state that model outputs were validated by UK clinical oncologists (CS B.3.10). No details are given about how this validation process was done or whether any changes were made as a result. It is also stated that the model was verified by economists not involved in its development. A list of verification checks is given, including checks on input data and technical validation of coding.

4.3.7.1 ERG model verification procedures

We conducted a range of manual checks to verify model inputs and calculations ('white box' tests) and to test the face-validity of the model results ('black box' checks):

- Cross-checking of all parameter inputs against values in the CS and cited sources;
- We traced input parameters from entry cells in the model ('User-Inputs" and "Resources" sheets), to PSA/DSA sampling (on the "Variables" sheet) through to the survival curve and Markov calculation sheets;
- We independently replicated calculations for first and second line drug costs (to check adjustments for dose, intensity, wastage and PAS discounts), health state costs and adverse event costs and QALY loss;
- Survival curve calculations were checked ("TPs_CABOSON", "TPs_ITC" and "TPs_ITC_FP" sheets).
- Use of PFS, OS and TTD results to estimate the distribution of the cohort by health state and the numbers of events over time in Markov trace sheets (E.Cabo.RCT etc.)
- We checked QALY and cost calculations on the Markov sheets;
- And the links from the total costs and outcomes on the Markov sheets back to the ICER calculations on the 'Results' sheet.
- We checked all model outputs against results cited in the CS, including the base case, PSA and DSA and we manually ran scenarios.

Through this process we identified some errors and inconsistencies:

1. QALY calculations – discounting and utility adjustment for age and sex were applied twice in the Markov trace sheets. This had the effect of shrinking the estimated QALYs for all treatments and hence the incremental QALY differences between treatments, thus overestimating the correct ICER.

2. QALY calculations – the one-off QALY loss that is applied in the first cycle to account for adverse effects of treatment was incorrectly adjusted for the duration of the cycle. The reduced the effect of adverse events on QALYs for all treatments.
3. The 'Accrual Utility' column in the Markov sheets also adjusted the QALYs accrued in each cycle again for the duration of the cycle. Thus the graphs of cumulative QALYs over time on the 'Table1' sheet are incorrect. This does not influence the cost-effectiveness results.
4. Utility estimates for PF and PD health states from the sunitinib NICE technology appraisal that were used in scenario analyses were incorrectly entered in the model. This was corrected by the company in response to a clarification question.
5. The cost of administering the nivolumab infusion was not included in calculation of second-line treatment costs. The unit cost for this administration cited in the model was also different for nivolumab than for bevacizumab and temsirolimus, which are also administered by infusion.
6. We believe that the cost of end of life care was incorrectly estimated for the NHS and PSS perspective and that it was incorrectly updated for inflation.
7. There was an error in the scenario analysis 'PFS=OS=Gompertz' (CS Table 66) (the PFS curve was set to lognormal rather than Gompertz).
8. The 'health resource (UK clinicians)' scenario analysis (CS Tables 66 and 67) gave the base case ICER because of an error in the linking of the Source of health resource control on the User_Input sheet.
9. On the 'Curve data' sheet, TTD for pazopanib was calculated as a proportion of pazopanib PFS, rather than being set equal to sunitinib TTD. This did not influence the cost-effectiveness results, although the summary statistics and graph for TTD on the 'User-inputs' sheet are incorrect.

4.3.7.2 Assessment of internal and external validity of model

Key statistics relating to the fit of the company's fitted survival models are shown in Table 13 and Table 14 for OS and PFS respectively. In addition to the model fit statistics (BIC and DIC), we show median survival and the proportion of the cohort progression free/ alive at 5 years for each fitted curve. For comparison, the tables include estimates of median and 5-year survival from other sources;

- the CABOSUN and COMPARZ trials.
- a cohort of patients from the IMDC database starting first-line treatment with sunitinib or pazopanib for metastatic RCC.[47]
- the committee's preferred model for the NICE appraisal of tivozanib (as reported in the published guidance).[22]

The COMPARZ and IMDC datasets and tivozanib model relate to patients with a mixed risk profile. Thus, estimates from these sources should be considered upper limits for survival for the intermediate/poor risk population in this appraisal.

Validity of fitted OS curves

For OS we have two sets of median survival estimates from CABOSUN, from the January 2017 and July 2017 cut-offs. The KM data used to fit the ITC models (clarification question B1) relates to the earlier cut-off. With respect to this dataset, several fitted models appear to underestimate median OS with cabozantinib, but for sunitinib median OS estimates were similar. For some of the fitted models, estimates of 5-year survival with sunitinib appear optimistic for the intermediate/poor risk population, as rates were similar in the IMDC cohort (21%).

Validity of fitted PFS curves

For sunitinib, most fitted models give estimates of median PFS that are similar to that in CABOSUN (5.3 months). Exceptions are the Weibull direct comparison; ITC exponential, Weibull and Gompertz; and first-order FPs. All the fitted models overestimated median PFS with cabozantinib with respect to CABOSUN (8.6 months). Median PFS estimates from the ITC models were slightly lower for pazopanib than for sunitinib; reflecting the small (but non-significant) PFS advantage for sunitinib in COMPARZ. As expected, all ITC models for sunitinib and pazopanib gave lower estimates of median PFS than the other sources. Five-year PFS with sunitinib and pazopanib was also lower for most fitted models than in the IMDC cohort, although sunitinib estimates from some FPs were similar to the 5% IMDC figure.

In summary, the ERG concludes that the company's preferred survival models (lognormal for the direct comparison and FP with P1=P2=-1 for the ITC) have reasonable face validity for sunitinib and pazopanib, with good measures of fit and similar median PFS as in the CABOSUN control arm. Both curves overestimate PFS for cabozantinib; yielding higher median PFS than in CABOSUN and a relatively large proportion of patients without disease progression at 5 years (5% and 10%). For OS, the company's preferred models (exponential direct comparison and FP with P1=P2=-1) also have a good fit for sunitinib and pazopanib and median survival is similar to that in the CABOSUN control arm (January 2017 data cut). For cabozantinib, the company's preferred models also give similar results to this dataset, but they overestimate median OS in relation to the most recent, July 2017 dataset. The plausibility of the company's survival extrapolations is unclear. 5-year survival with cabozantinib is estimated at 21% and 24% with the company's preferred direct and ITC models. To put this in

perspective, this is similar to 5-year OS in the IMDC cohort, who had a more favourable risk profile but were treated at first line with sunitinib or pazopanib.

Although we are critical of the apparent overestimation of PFS and OS for cabozantinib with the company's preferred methods, the other fitted models do not address these concerns.

4.3.8 Cost effectiveness results

Results from the company's economic model are presented in section B.3.7 of the CS. The ERG believes that these results include some errors in model inputs and calculations, as described in the previous section. For comparison, we reproduce the CS original results in Table 24 below and ERG corrected results in section 4.4.2.

For the company base case using the direct comparison from CABOSUN, an ICER of £37,793 per QALY gained is reported for cabozantinib versus sunitinib. Based on the company's preferred ITC model, sunitinib is dominated by pazopanib and the ICER for cabozantinib compared with pazopanib is £48,451. The pairwise ICER for cabozantinib compared with sunitinib in this model is £31,538.

Table 24 Company base-case results, deterministic (from CS Tables 60 and 61)

Drug	Costs (£)	QALYs	Life-years	ICER (£ per QALY gained)	
				Incremental analysis	Pairwise, cabozantinib vs. comparator
Direct comparison (CABOSUN)					
Sunitinib	[redacted]	[redacted]	[redacted]	-	-
Cabozantinib	[redacted]	[redacted]	[redacted]	37,793	37,793
ITC comparison (CABOSUN and COMPARZ)					
Pazopanib	[redacted]	[redacted]	[redacted]	-	48,451
Sunitinib	[redacted]	[redacted]	[redacted]	Dominated	31,538
Cabozantinib	[redacted]	[redacted]	[redacted]	48,451	-

The CS states that cabozantinib is an effective treatment for advanced RCC in treatment naïve patients when compared with sunitinib and pazopanib. No claims are made regarding cost-effectiveness and the company did not carry out any economic analysis for subgroups. The company's approach to handling uncertainty is discussed below.

4.3.9 Assessment of uncertainty

4.3.9.1 Probabilistic sensitivity analysis

The company PSA results are summarised in scatterplots, CEACs and tables of incremental cost per QALY gained (CS Figures 17 to 22: CS Tables 62 to 64). The PSA results are stable and similar to the deterministic results. The CS summarises the probabilistic results stating that there is a 66.1% probability (based on the CABOSUN study) or a 74.4% probability (based on the ITC result) of cabozantinib being cost-effective, relative to sunitinib, at a threshold willingness to pay of £50,000 per QALY gained. Relative to pazopanib, the CS quotes a 47.8% probability (based on the ITC result) of cabozantinib being cost-effective, at a threshold willingness to pay of £50,000 per QALY gained.

4.3.9.2 One-way sensitivity analyses

One-way sensitivity analyses are undertaken and reported in the CS. Model parameters are varied across a range to test the sensitivity of the ICERs to individual parameters or groups of parameters. The CS reports the input ranges and distributions for the model parameters in CS Table 65. The results are summarised in the tornado graphs in CS Figures 23 and 24. The company does not expressly justify the ranges used for the one-way sensitivity analysis. However, and most parameter ranges are based on observed values, such as 95% CI, and choice of distributions (CS Table 65) is reasonable. The tornado graphs only show parameters that make at least £1000 per QALY gained difference between the minimum and maximum limits. These include drug costs and discount rates for QALYs and costs, which have the biggest impact on cost-effectiveness. Other than these parameters, the key drivers of cost-effectiveness are relative dose intensity and utilities associated with the progression free state. However, we note that this analysis does not reflect uncertainties over the treatment effects on PFS, OS and TTD: structural uncertainty over the choice of survival curve analysis method; or uncertainty around the fitted parameters for those curves. The impact of these uncertainties is reflected in the PSA and scenario analyses.

4.3.9.3 Scenario analyses

The company explores a range of scenarios which are reported in the CS Table 66 and 67 Some of the company's scenario analyses were informed by expert opinion. Generally, the company appears to test scenarios using available data that was not used in the base case. The company found that the biggest source of uncertainty over cost-effectiveness was the choice of OS curve used in the model.

4.4 Additional work undertaken by the ERG

4.4.1 Description and justification of ERG analyses

Table 25 shows the corrections the ERG made to the company's model. Table 26 shows our preferred assumptions and scenarios, and Table 27 shows our approach to modelling treatment effects.

Table 25 ERG corrections to company model

Aspect of model	Problem	ERG correction
QALY calculations	1. Discounting and adjustment of utilities for age and sex are applied twice. 2. QALY loss for adverse events applied at first cycle is incorrectly adjusted for the duration of the cycle. 3. Accrual utility adjusted again for duration of cycle.	Columns AA to AF on Markov trace sheets ('E.Cabo.RCT' etc.) recoded
Health state utilities for scenario	4. Incorrect values for sunitinib TA scenario analysis in 'Resources' sheet (cells F231-M231).	Corrected in company response to clarification question.
Administration cost for nivolumab	5. Cost for administration not included in cost calculation. The cost cited on the User-Inputs page also differs to that for bevacizumab and temsirolimus.	Changed admin cost to £205 (cell I126 'User_Inputs') and added to weekly cost calculation for nivolumab (cell F97 'Variables').
Cost of end of life care at last cycle before death	6. Cost used in company base case only relates to hospital care. Costs for local authority funded social care, district nursing and GP visits excluded. Uprated using general price inflation (not health specific).	ERG estimated value of £7,961 for NHS and PSS perspective uprated from 2010/11 to 2016/17 using HCHS index.[52]
Scenario analyses	7. Scenario with PFS=OS=Gompertz used lognormal rather than Gompertz distribution for PFS. 8. Health resource (UK clinicians) scenario gives base case result.	'ScenarioAnalysis' cell N33 changed to 5 (Gompertz). Control in cell F154 on the 'User_Inputs' sheet linked to v.vHealthResource.Input.
TTD on curve data sheet and graph	9. TTD curve for pazopanib defined in relation to pazopanib PFS curve. Gives wrong summary statistics and TTD curve on 'User_Inputs' sheet.	Deleted 'Curve data' sheet and replaced summary statistic calculations and figures linked to 'TPs_CABOSUN' etc. sheets

Table 26 ERG preferred assumptions and scenarios

	Preferred assumptions	Scenarios	Reason for analysis
Time horizon	20 years	5/ 10 years	Reflects full lifetime, but with scenario analysis to show impact of extrapolation
Persistence of OS and PFS benefit	5 years from baseline	10/ 20 years	Given the weakness of evidence for the OS difference, we take a conservative approach, with progression and mortality hazards for cabozantinib equal to those of sunitinib after 5 years (3 years after trial follow up).
OS curves	Simple indirect comparison	HR = 0.74 (Jan 2017 analysis). And no effect (HR=1)	Exponential OS for sunitinib (separate fit to CABOSUN). Cabozantinib estimated from sunitinib curve and HR=80 (July 2017 CABOSUN update). OS assumed equal for pazopanib and sunitinib, based on COMPARZ. Exploratory scenarios to compare with company model and assess impact of OS.
	Age-related mortality		Minimum mortality rate based on general population life table (ONS 2014-16).
PFS curves	Lognormal direct comparison	Exponential and Gompertz	Same as in company direct base case. Lognormal gives most plausible fit, and we use selected alternatives for scenarios (see table below).
TTD curves	Lognormal direct comparison	All available	We agree that the lognormal gives the best fit, but there is little reason to choose between other functions, so we use all in scenario analyses.
Health state utilities	PF and PD utilities from Tivozanib TA512 (base case)	Swinburn, Pazo TA215 and Suni TA169	We follow the company approach, with the utilities for pre and post-progression based on values accepted by committee for tivozanib, with scenarios testing alternative sources.
AE disutilities	Amdahl disutility, applied for 4 weeks to TEAE with >=5% incidence	Range of disutilites, 8 week duration and >=2%	Again, we follow the company approach, but conduct additional analyses to test the sensitivity of the model to adverse events.

	Preferred assumptions	Scenarios	Reason for analysis
Dose intensities	Dose intensities from CABOSUN (94.3% cabo, 83.9% suni) and 86% for pazo from tivozanib STA	Tested 86% for all first-line drugs, and also 100%	Company's assumptions are reasonable but we explore the impact on costs of uncertainty over dose intensity, using the range suggested by committee considerations from the NICE tivozanib appraisal guidance
Subsequent treatment costs	Use of second-line treatments from trials	Company and ERG scenarios	Utilisation from trials reflects effectiveness evidence, but it includes drugs not recommended or available in UK. The company includes a scenario based on clinical advice, using only NICE recommended second-line drugs. We test 2 other scenarios. ERG 1: equal distribution of NICE approved second-line drugs (20% each drug and 10% BSC; cabozantinib 1st line patients only eligible for nivolumab, everolimus or lenvatinib with everolimus, 30% each drug and 10% BSC). ERG 2: based on clinical advice we assume use only of nivolumab, cabozantinib, lenvatinib with everolimus (30% each drug, and 10% BSC; cabozantinib 1st line patients only eligible for nivolumab and lenvatinib with everolimus, 45% each drug and 10% BSC).
Health state management costs	Based on resource use assumptions from tivozanib appraisal	Company scenario based on clinical advice. More expensive blood test (£20)	Clinical advisors to the ERG agreed that resource use assumptions were appropriate
Adverse event costs	Series of assumptions based on clinical advice and guidance.		As above
Age of cohort	years	55/75 years	Exploratory: to assess applicability to the UK RCC population

Table 27 ERG approach to modelling treatment effects

	Company base case (scenarios)	Comments	ERG preferred assumptions
OS curves	**Direct**: Exponential (Weibull & Gompertz) **ITC**: FP model with P1=P2=-1 (exponential; Weibull; Gompertz; and FP P1=-0.5, P2=0 & P1=-1, P2=0)	CABOSUN is not powered for OS and data are relatively immature, so the KM curves are noisy. Reason for crossover is unclear. Uncertainties over the ITCs due to differences in trial populations. Given these reservations, the exponential, Weibull and Gompertz are reasonable for the direct analysis. For the ITC, the exponential and FP P1=P2=-1 curves are reasonable. But other scenarios predict unrealistic long-term survival. Fitted curves based on Jan 2017 CABOSUN data, rather than less favourable July 2017 dataset.	Simple indirect comparison assuming: • Sunitinib OS curve based on company's exponential fit to CABOSUN; • Cabozantinib calculated from sunitinib curve and HR form July 2017 CABOSUN results; • Pazopanib curve assumed equal to sunitinib (based on COMPARZ results).
PFS curves	**Direct**: lognormal (Exponential, Weibull & Gompertz) **ITC**: FP P1=P2=-1 (exponential, Weibull and Gompertz)	CABOSUN PFS analysis is more mature. ITC is subject to uncertainty due to differences in trial populations, unclear if similarity assumption is met. Direct comparisons with lognormal, exponential and Gompertz are reasonable, but the Weibull has poor visual fit. For ITC, Lognormal and loglogistic models give best balance of fit and extrapolation.	Simple indirect comparison: use lognormal separately fitted to CABOSUN for cabozantinib and sunitinib and assume equivalence for pazopanib and sunitinib (COMPARZ). We also test alternative separately fitted curves: exponential and Gompertz curves.
TTD curves	**Direct**: lognormal (exponential, Weibull, Gompertz & gamma).	TTD data are mature, with little difference in the visual fit or extrapolation of survival functions. There is no obvious reason for excluding the loglogistic from scenario analysis. The assumption of equal TTD for pazopanib and sunitinib is reasonable given similarity in COMPARZ.	Lognormal for base case, and all other distributions in scenario analysis.

4.4.2 Results of ERG analyses

All analyses in this report reflect agreed PAS discounts for cabozantinib and pazopanib, and the free first cycle for sunitinib. However, they exclude PAS discounts for subsequent treatments. PAS discounts are in place for cabozantinib, axitinib, nivolumab, everolimus and lenvatinib. We replicate the tables below with PAS discounts in a confidential addendum to this report.

4.4.2.1 ERG corrections to company analyses

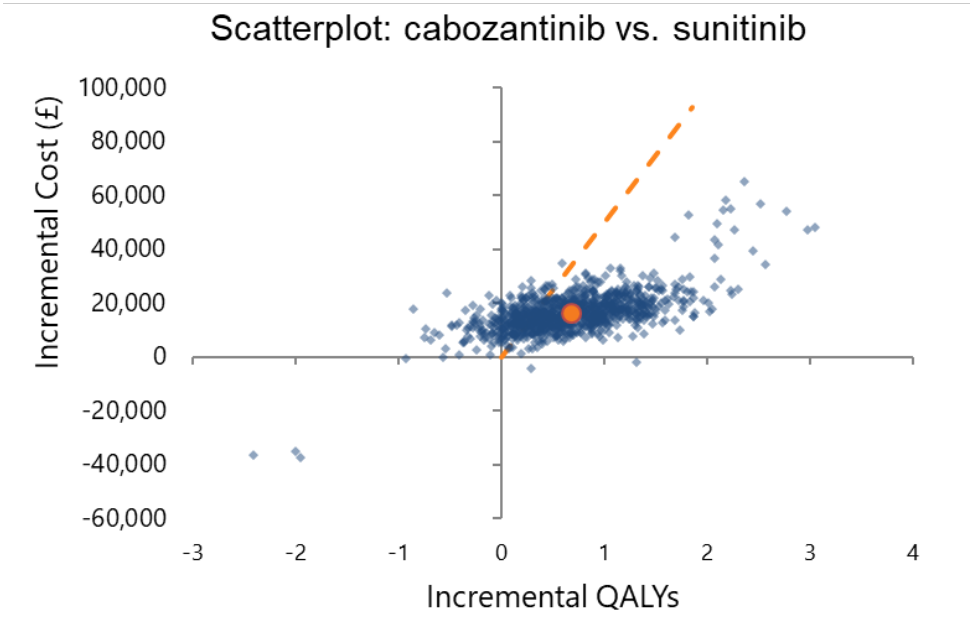
Corrections to the QALY calculations (points 1-3 in Table 25) increase the QALY estimates for all treatments. This increases the incremental QALY gains for cabozantinib, hence reducing ICERs. For example, based on the direct comparison with sunitinib, incremental QALYs increase from 0.401 in the original company base case to 0.471 in our corrected analysis, which reduces the ICER from £37,392 to £32,340 per QALY gained. Corrections to the costs of nivolumab and end of life care further reduce the ICER estimates: e.g. to £31,956 for the direct comparison of cabozantinib with sunitinib. The cost-effectiveness results from the ERG corrections to the company's base case analyses are shown in Table 28. These show that sunitinib is dominated by pazopanib, which yields more QALYs at a lower cost. The ICER for cabozantinib compared with pazopanib is £40,757 per QALY gained. Compared with sunitinib, cabozantinib has an ICER of £31,956 per QALY gained based on the direct comparison from CABOSUN data, and £26,182 per QALY gained based on the company's preferred indirect comparison using CABOSUN and COMPARZ data.

Table 28 Cost-effectiveness: Company base-case analyses (ERG corrected)

Drug	Costs (£)	QALYs	Life-years	ICER (£ per QALY gained)	
				Incremental analysis	Pairwise, cabozantinib vs. comparator
Direct comparison (CABOSUN)					
Sunitinib	[redacted]	[redacted]	[redacted]	-	31,956
Cabozantinib	[redacted]	[redacted]	[redacted]	31,956	-
ITC (CABOSUN and COMPARZ)					
Pazopanib	[redacted]	[redacted]	[redacted]	-	40,757
Sunitinib	[redacted]	[redacted]	[redacted]	Dominated	26,182
Cabozantinib	[redacted]	[redacted]	[redacted]	40,757	-

Results from the probabilistic sensitivity analysis (PSA) are similar. The extent of uncertainty around the incremental costs and QALYs for cabozantinib compared with sunitinib and pazopanib is illustrated in the scatterplots in Figure 20 (for the company's ITC base case).

Based on 1,000 PSA iterations, there is an estimated probability that cabozantinib is cost-effective compared with pazopanib is 28% at a cost effectiveness threshold of £30,000 per QALY gained and 57% at a threshold of £50,000 per QALY gained.



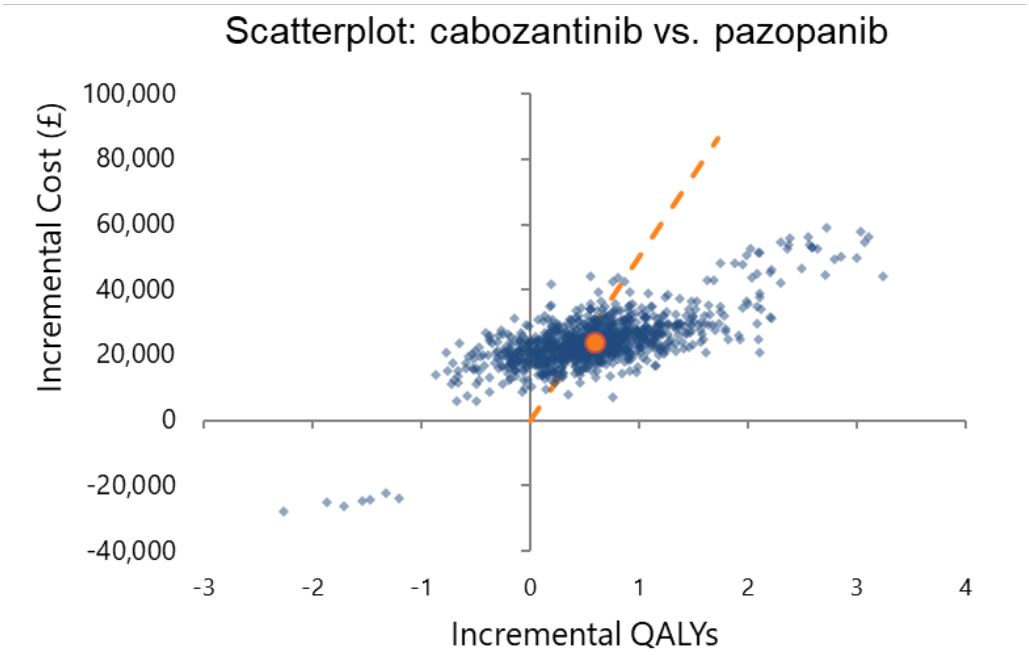


Figure 20 CE scatterplots, company ITC base case (ERG corrected)

Results from the one-way deterministic sensitivity analyses on the ERG-corrected version of the company's ITC base case are illustrated in Figure 21 and Figure 22. These suggest that the cost and relative dose intensity of the treatment and comparator as well as other cost parameters are the key drivers of cost-effectiveness. However, this is misleading, as effectiveness parameters are not included in this analysis.

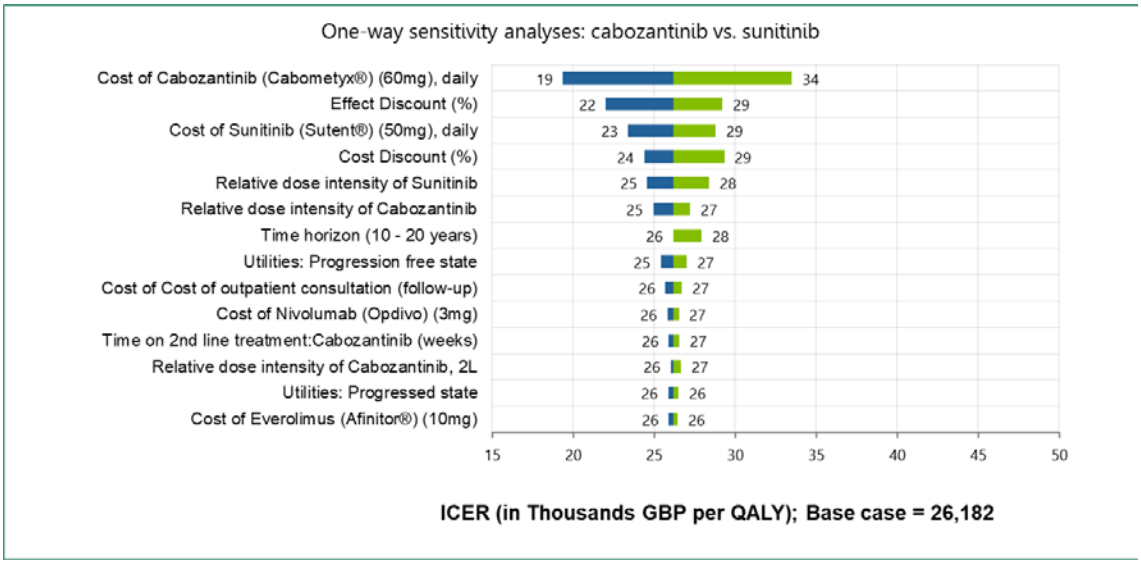


Figure 21 Tornado diagram: Company ITC base case (ERG corrected)

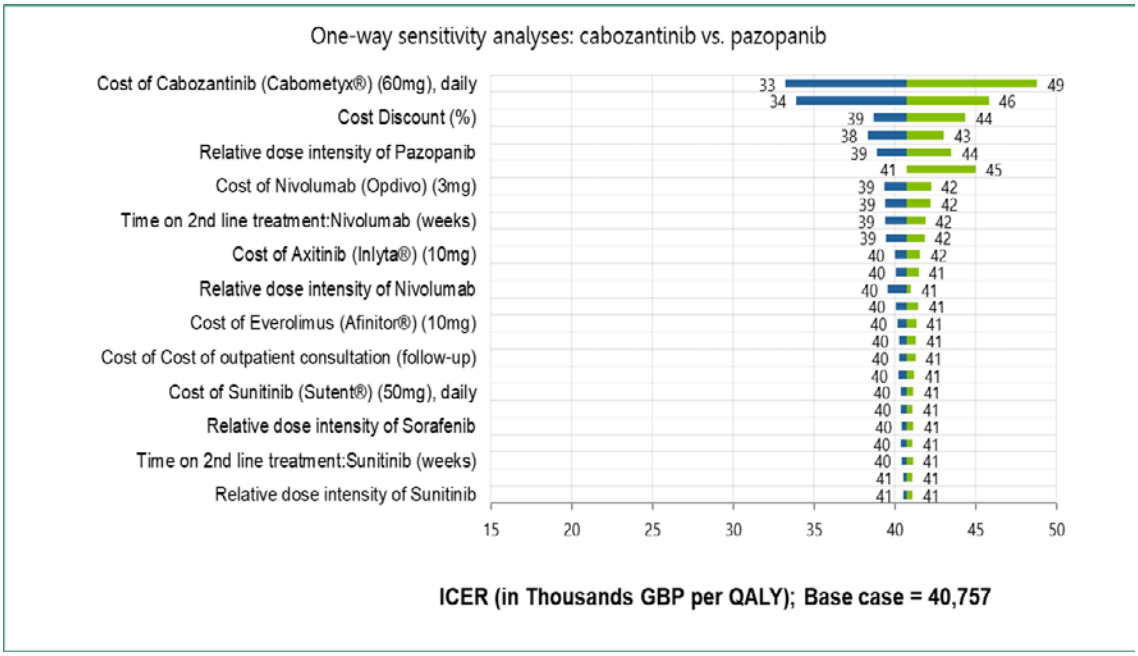


Figure 22 Tornado diagram: Company ITC base case (ERG corrected)

The impact of key uncertainties over model assumptions and data sources of data is shown in Table 29, Table 30 and Table 31 below. The model is most sensitive to assumptions and methods of fitting the OS curves. The model is also sensitive to a very short time horizon.

Table 29 Scenario analysis: Company direct base case (ERG corrected) vs. sunitinib

Scenario		Total cost (£)		Total QALYs		ICER (£ per QALY)
		Cabo	Suni	Cabo	Suni	
Company direct base case		[redacted]	[redacted]	[redacted]	[redacted]	**31,956**
Time horizon	10 years	[redacted]	[redacted]	[redacted]	[redacted]	33,216
	5 years	[redacted]	[redacted]	[redacted]	[redacted]	40,719
PFS curves	Exponential	[redacted]	[redacted]	[redacted]	[redacted]	30,414
	Weibull	[redacted]	[redacted]	[redacted]	[redacted]	29,247
	Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	31,562
	Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	31,749
	Gamma	[redacted]	[redacted]	[redacted]	[redacted]	30,671
OS curves	Weibull	[redacted]	[redacted]	[redacted]	[redacted]	41,669
	Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	30,226
	Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	38,946
	Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	47,576
	Gamma	[redacted]	[redacted]	[redacted]	[redacted]	20,841
TTD curves	Exponential	[redacted]	[redacted]	[redacted]	[redacted]	26,586
	Weibull	[redacted]	[redacted]	[redacted]	[redacted]	26,596
	Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	28,978
	Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	33,022
	Gamma	[redacted]	[redacted]	[redacted]	[redacted]	29,879
Utility source	PF and PD (Swinburn)	[redacted]	[redacted]	[redacted]	[redacted]	27,461
	PF and PD (Pazo TA215)	[redacted]	[redacted]	[redacted]	[redacted]	32,912
	PF and PD (Suni TA169)	[redacted]	[redacted]	[redacted]	[redacted]	29,779
	TEAE (METEOR)	[redacted]	[redacted]	[redacted]	[redacted]	31,893
Costs	Comprehensive blood test	[redacted]	[redacted]	[redacted]	[redacted]	32,266
	Management (UK clinician)	[redacted]	[redacted]	[redacted]	[redacted]	30,595
	Second-line (UK practice)	[redacted]	[redacted]	[redacted]	[redacted]	34,081
	No infusion wastage	[redacted]	[redacted]	[redacted]	[redacted]	32,099
	No end of life cost	[redacted]	[redacted]	[redacted]	[redacted]	32,349

Table 30 Scenario analysis: Company ITC base case (ERG corrected), vs. sunitinib

Scenario		Total cost (£)		Total QALYs		ICER (£ per QALY)
		Cabo	Suni	Cabo	Suni	
Company ITC base case		[redacted]	[redacted]	[redacted]	[redacted]	**26,182**
Time horizon	10 years	[redacted]	[redacted]	[redacted]	[redacted]	27,912
	5 years	[redacted]	[redacted]	[redacted]	[redacted]	35,488
PFS curves	FP P1=-0.5 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	25,795
	FP P1=-1 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	25,818
	ITC RE exponential	[redacted]	[redacted]	[redacted]	[redacted]	28,909
	ITC RE Weibull	[redacted]	[redacted]	[redacted]	[redacted]	28,551
	ITC RE Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	29,043
	ITC RE Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	30,094
	ITC RE Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	29,700
OS curves	FP P1=-0.5 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	55,215
	FP P1=-1 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	33,356
	ITC RE Exponential	[redacted]	[redacted]	[redacted]	[redacted]	30,094
	ITC RE Weibull	[redacted]	[redacted]	[redacted]	[redacted]	38,252
	ITC RE Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	23,445
	ITC RE Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	45,415
	ITC RE Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	49,983
TTD curves	Exponential	[redacted]	[redacted]	[redacted]	[redacted]	21,816
	Weibull	[redacted]	[redacted]	[redacted]	[redacted]	21,826
	Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	23,760
	Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	27,702
	Gamma	[redacted]	[redacted]	[redacted]	[redacted]	24,475
Utility source	PF and PD (Swinburn)	[redacted]	[redacted]	[redacted]	[redacted]	21,332
	PF and PD (Pazo TA215)	[redacted]	[redacted]	[redacted]	[redacted]	26,787
	PF and PD (Suni TA169)	[redacted]	[redacted]	[redacted]	[redacted]	24,431
	TEAE (METEOR)	[redacted]	[redacted]	[redacted]	[redacted]	26,141
Costs	Comprehensive blood test	[redacted]	[redacted]	[redacted]	[redacted]	26,488
	Management (UK clinician)	[redacted]	[redacted]	[redacted]	[redacted]	24,926
	Second-line (UK practice)	[redacted]	[redacted]	[redacted]	[redacted]	28,425
	No infusion wastage	[redacted]	[redacted]	[redacted]	[redacted]	26,262
	No end of life cost	[redacted]	[redacted]	[redacted]	[redacted]	26,585

Table 31 Scenario analysis: Company ITC base case (ERG corrected), vs. pazopanib

Scenario		Total cost (£)		Total QALYs		ICER (£ per QALY)
		Cabo	Pazo	Cabo	Pazo	
Company ITC base case		[redacted]	[redacted]	[redacted]	[redacted]	**40,757**
Time horizon	10 years	[redacted]	[redacted]	[redacted]	[redacted]	45,001
	5 years	[redacted]	[redacted]	[redacted]	[redacted]	64,841
PFS curves	FP P1=-0.5 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	39,653
	FP P1=-1 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	39,733
	ITC RE exponential	[redacted]	[redacted]	[redacted]	[redacted]	50,540
	ITC RE Weibull	[redacted]	[redacted]	[redacted]	[redacted]	50,591
	ITC RE Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	49,206
	ITC RE Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	51,910
	ITC RE Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	50,037
OS curves	FP P1=-0.5 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	74,858
	FP P1=-1 P2=0	[redacted]	[redacted]	[redacted]	[redacted]	49,973
	ITC RE Exponential	[redacted]	[redacted]	[redacted]	[redacted]	51,910
	ITC RE Weibull	[redacted]	[redacted]	[redacted]	[redacted]	65,942
	ITC RE Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	37,788
	ITC RE Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	78,883
	ITC RE Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	86,300
TTD curves	Exponential	[redacted]	[redacted]	[redacted]	[redacted]	36,236
	Weibull	[redacted]	[redacted]	[redacted]	[redacted]	36,188
	Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	38,277
	Loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	42,564
	Gamma	[redacted]	[redacted]	[redacted]	[redacted]	38,826
Utility source	PF and PD (Swinburn)	[redacted]	[redacted]	[redacted]	[redacted]	31,471
	PF and PD (Pazo TA215)	[redacted]	[redacted]	[redacted]	[redacted]	41,419
	PF and PD (Suni TA169)	[redacted]	[redacted]	[redacted]	[redacted]	38,073
	TEAE (METEOR)	[redacted]	[redacted]	[redacted]	[redacted]	40,578
Costs	Comprehensive blood test	[redacted]	[redacted]	[redacted]	[redacted]	41,060
	Management (UK clinician)	[redacted]	[redacted]	[redacted]	[redacted]	39,603
	Second-line (UK practice)	[redacted]	[redacted]	[redacted]	[redacted]	26,736
	No infusion wastage	[redacted]	[redacted]	[redacted]	[redacted]	39,979
	No end of life cost	[redacted]	[redacted]	[redacted]	[redacted]	41,159

4.4.2.2 ERG preferred analysis

Results based on the ERG preferred assumptions are shown in Table 32. As in the company analyses, sunitinib is dominated as pazopanib is less expensive and no less effective. Compared with pazopanib, cabozantinib has an ICER of £65,743 per QALY gained. The ICER for cabozantinib is £41,465 in comparison with sunitinib. By assumption, life expectancy is the same for pazopanib and sunitinib in this analysis and there is a small difference in mean progression-free life years between these comparators. Cabozantinib has a modest survival advantage and a larger effect on progression free-survival. We believe that these estimates appropriately reflect the evidence from the CABOSUN and COMPARZ trials.

Table 32 Cost-effectiveness: ERG preferred assumptions

Drug	Costs (£)	QALYs	Life-years	PF life years	ICER (£ per QALY gained)	
					Incremental analysis	Pairwise, cabozantinib vs. comparator
Pazopanib	[redacted]	[redacted]	[redacted]	[redacted]	-	65,743
Sunitinib	[redacted]	[redacted]	[redacted]	[redacted]	-	41,465
Cabozantinib	[redacted]	[redacted]	[redacted]	[redacted]	65,743	-

Table 33 and Table 34 summarise scenario analyses around our preferred set of assumptions for the comparison of cabozantinib versus pazopanib and sunitinib respectively. Generally, the results are robust, with the pairwise ICERs remaining above £30,000 per QALY gained for all scenarios tested. The ICERs were most sensitive to the assumption that cabozantinib has no relative effect on survival compared with sunitinib or pazopanib. This illustrates that the results are very largely driven by the effect on OS, as estimated from the CABOSUN trial.

Table 33 Scenario analysis: ERG preferred assumptions, vs. pazopanib

Scenario		Total cost (£)		Total QALY		ICER (£)
		Cabo.	Pazo.	Cabo.	Pazo.	
ERG preferred assumptions		[redacted]	[redacted]	[redacted]	[redacted]	**65,743**
Time horizon	5 years	[redacted]	[redacted]	[redacted]	[redacted]	79,127
	10 years	[redacted]	[redacted]	[redacted]	[redacted]	66,783
Persistence of OS/ PFS effect	10 years	[redacted]	[redacted]	[redacted]	[redacted]	58,890
	20 years	[redacted]	[redacted]	[redacted]	[redacted]	57,879
CABOSUN OS curves	HR = 0.74 (Jan 2017)	[redacted]	[redacted]	[redacted]	[redacted]	52,778
	No effect on OS	[redacted]	[redacted]	[redacted]	[redacted]	372,866
CABOSUN PFS curves	Separate exponential	[redacted]	[redacted]	[redacted]	[redacted]	64,913
	Separate Weibull	[redacted]	[redacted]	[redacted]	[redacted]	64,192
	Separate Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	64,880
TTD curves	Separate exponential	[redacted]	[redacted]	[redacted]	[redacted]	59,908
	Separate Weibull	[redacted]	[redacted]	[redacted]	[redacted]	59,836
	Separate Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	63,012
	Separate loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	65,638
	Separate gamma	[redacted]	[redacted]	[redacted]	[redacted]	64,092
Utility values	Swinburn	[redacted]	[redacted]	[redacted]	[redacted]	47,616
	Pazo NICE STA	[redacted]	[redacted]	[redacted]	[redacted]	66,246
	Suni NICE STA	[redacted]	[redacted]	[redacted]	[redacted]	61,500
TEAE disutility	METEOR (-0.05)	[redacted]	[redacted]	[redacted]	[redacted]	65,224
	Higher disutility (-0.4)	[redacted]	[redacted]	[redacted]	[redacted]	66,436
	Include if >= 2%	[redacted]	[redacted]	[redacted]	[redacted]	64,863
	Duration: 8 weeks	[redacted]	[redacted]	[redacted]	[redacted]	66,468
Drug costs (first line)	Dose intensities 86%	[redacted]	[redacted]	[redacted]	[redacted]	58,517
	Does intensities 100%	[redacted]	[redacted]	[redacted]	[redacted]	65,739
Drug costs (second line)	% use (Company)	[redacted]	[redacted]	[redacted]	[redacted]	41,936
	% use (ERG 1)	[redacted]	[redacted]	[redacted]	[redacted]	45,980
	% use (ERG 2)	[redacted]	[redacted]	[redacted]	[redacted]	44,374
Other costs	Blood test (£20)	[redacted]	[redacted]	[redacted]	[redacted]	66,039
	Follow up (UK clinician)	[redacted]	[redacted]	[redacted]	[redacted]	64,738
	No end of life cost	[redacted]	[redacted]	[redacted]	[redacted]	66,106
Age of cohort	55 years	[redacted]	[redacted]	[redacted]	[redacted]	65,567
	75 years	[redacted]	[redacted]	[redacted]	[redacted]	66,061

Table 34 Scenario analysis: ERG preferred assumptions, vs. sunitinib

Scenario		Total cost (£)		Total QALY		ICER (£)
		Cabo.	Suni.	Cabo.	Suni.	
ERG preferred assumptions		[redacted]	[redacted]	[redacted]	[redacted]	**41,465**
Time horizon	5 years	[redacted]	[redacted]	[redacted]	[redacted]	46,564
	10 years	[redacted]	[redacted]	[redacted]	[redacted]	41,839
Persistence of OS/ PFS effect	10 years	[redacted]	[redacted]	[redacted]	[redacted]	37,716
	20 years	[redacted]	[redacted]	[redacted]	[redacted]	37,170
CABOSUN OS curves	Exponential, HR 0.74	[redacted]	[redacted]	[redacted]	[redacted]	34,202
	No effect on OS	[redacted]	[redacted]	[redacted]	[redacted]	204,789
CABOSUN PFS curves	Separate exponential	[redacted]	[redacted]	[redacted]	[redacted]	39,904
	Separate Weibull	[redacted]	[redacted]	[redacted]	[redacted]	38,871
	Separate Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	40,107
TTD curves	Separate exponential	[redacted]	[redacted]	[redacted]	[redacted]	35,219
	Separate Weibull	[redacted]	[redacted]	[redacted]	[redacted]	35,237
	Separate Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	38,267
	Separate loglogistic	[redacted]	[redacted]	[redacted]	[redacted]	41,428
	Separate gamma	[redacted]	[redacted]	[redacted]	[redacted]	39,696
Utility values	Swinburn	[redacted]	[redacted]	[redacted]	[redacted]	30,089
	Pazo NICE STA	[redacted]	[redacted]	[redacted]	[redacted]	41,780
	Suni NICE STA	[redacted]	[redacted]	[redacted]	[redacted]	38,805
TEAE disutility	METEOR (-0.05)	[redacted]	[redacted]	[redacted]	[redacted]	41,346
	Higher disutility (-0.4)	[redacted]	[redacted]	[redacted]	[redacted]	41,621
	Include if >= 2%	[redacted]	[redacted]	[redacted]	[redacted]	41,026
	Duration: 8 weeks	[redacted]	[redacted]	[redacted]	[redacted]	41,628
Drug costs (first line)	Dose intensities 86%	[redacted]	[redacted]	[redacted]	[redacted]	34,713
	Does intensities 100%	[redacted]	[redacted]	[redacted]	[redacted]	42,158
Drug costs (second line)	% use (Company)	[redacted]	[redacted]	[redacted]	[redacted]	43,856
	% use (ERG 1)	[redacted]	[redacted]	[redacted]	[redacted]	47,872
	% use (ERG 2)	[redacted]	[redacted]	[redacted]	[redacted]	46,276
Other costs	Blood test (£20)	[redacted]	[redacted]	[redacted]	[redacted]	41,759
	Follow up (UK clinician)	[redacted]	[redacted]	[redacted]	[redacted]	40,466
	No end of life cost	[redacted]	[redacted]	[redacted]	[redacted]	41,825
Age of cohort	55 years	[redacted]	[redacted]	[redacted]	[redacted]	41,354
	75 years	[redacted]	[redacted]	[redacted]	[redacted]	41,664

5 End of life

The CS argues that cabozantinib meets the NICE end-of-life criteria. Table 35 (CS Table 28) summarises their justification for reaching this conclusion.

Table 35 End-of-life criteria (CS Table 28)

Criterion	Data available
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	In the IMDC validation study (1028 patients receiving first line VEGF-targeted treatment for metastatic RCC), median OS from the start of treatment was 22.5 months (18.7-25.1) in the intermediate risk group and 7.8 months (6.5-9.7) in the poor risk group.
There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment	In the CABOSUN trial, median survival was 30.3 months (95% CI 14.6, NE) in the cabozantinib arm vs. 21.0 months (95% CI 16.3, 27.0) in the sunitinib arm, an estimated 9.3 -month difference in the medians at a median follow-up of 28.9 months. In the economic modelling, which extrapolates beyond the duration of the trial, cabozantinib is associated with a gain of 0.66 life years (7.9 months) compared with sunitinib. The other treatment currently used in the NHS is pazopanib. Pazopanib was found to have similar efficacy to sunitinib in terms of both PFS and OS in a head-to-head trial in 1110 patients with previously untreated metastatic RCC (Motzer 2013). In the economic modelling, cabozantinib is associated with a gain of 0.80 life years (9.6 months) compared with pazopanib.

The ERG's analysis confirms that cabozantinib offers an additional extension of life, which exceeds 3 months when compared to sunitinib or pazopanib (5.9 months in ERG's analysis). However, the submitted CS model and results from the ERG's preferred assumptions give mean OS estimates exceeding 24 months for sunitinib and pazopanib ([redacted] life years without discounting in the ERG analysis). We are therefore of the opinion that cabozantinib does not fully meet the NICE criteria for being considered as a life-extending treatment for people with a short life expectancy.

6 Innovation

The CS suggests that the superior effectiveness compared with current treatments can be explained by its novel mechanism of action. Cabozantinib is the first and only multi-targeted therapy for RCC which targets pathways involved in both tumour growth and drug resistance

(MET, AXL), as well as tumour angiogenesis (VEGF). It is stated that by targeting MET and AXL receptors in addition to VEGFR, cabozantinib may provide additional anticancer efficacy over the more selective, existing anti-VEGFR agents (B.2.12).

Expert clinical advice to the ERG suggests that all of the currently available drugs inhibit VEGF, which is thought to be the main mechanism of action in RCC. Cabozantinib is not the only drug therapy that targets other pathways. However, it is not yet clear how important these other pathways are in drug efficacy.

The ERG notes that in the previous NICE appraisal of cabozantinib for previously treated RCC (TA 463)[14] it was accepted that cabozantinib would likely have additional benefits for some patients due to its multi-targeted approach, and could therefore be considered innovative. However, cabozantinib was not considered to reflect a 'step change' in treatment (The ERG infers that this consideration is within the context of previously treated RCC patients, not necessarily within the context of untreated RCC).

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The results of the CABOSUN trial show a statistically significant effect on PFS, the primary outcome, with a median PFS of 8.6 months (95% CI 6.8, 14.0) for cabozantinib and 5.3 months (95% CI 3.0, 8.2) for sunitinib (p=0.0008). The median difference of 3.3 months favoured cabozantinib. The confidence intervals around the PFS estimates are reasonably narrow indicating greater certainty in the estimates. It is important to put these results into context of the results of other trials of first line drug therapies in RCC. Rini and Vogelzang[63] discussed the results of the CABOSUN trial and noted that the median PFS of 5.6 months for the sunitinib arm of the CABOSUN trial was lower than that achieved in previous clinical trials. Specifically, in the phase III registration trial for sunitinib,[64] the median PFS for patients in the intermediate risk subgroup was 10.6 months. The ERG notes that this trial had a slightly lower percentage of patients with bone metastases and lower percentage of patients with prior nephrectomy than CABOSUN, which suggests slightly more favourable prognostic characteristics. Nonetheless, it can be considered an informative benchmark for PFS. The ERG notes that in a recently published phase III RCT comparing nivolumab in combination with ipilimumab versus sunitinib

in previously untreated clear-cell advanced RCC, median PFS for the sunitinib arm in the intermediate/poor risk subgroup was 8.4 months. Of note, the statistical power calculation in the CABOSUN trial assumed a median PFS of 8 months for sunitinib. This is 2.7 months higher than the median PFS achieved. The CS does not comment on this.

The CS cites a registry study of 1189 previously untreated poor and intermediate risk patients receiving targeted therapies (among whom sunitinib was the most common treatment), which reported a PFS of 5.6 months.[65] The CS suggests this is consistent with the CABOSUN results. However, Rini and Vogelzang[63] note that this data set included patients with non–clear cell histology (12%), patients with sarcomatoid histology (10%), and patients who received sorafenib, temsirolimus, or everolimus (21%). They suggest that these features might be expected to result in a lower PFS than would be expected in practice and the benchmark of 5.6 months isn't necessarily realistic.

Choueiri et al[66] (the CABOSUN trial investigators) responded to Rini and Vogelzang[63] that the CABOSUN trial included patients with high rates of poor prognostic clinical factors, which distinguishes it from other contemporary trials of untreated patients with metastatic RCC. They note that PFS was also shorter in a retrospective community setting study of sunitinib (7.5 months) in 134 patients.[67] They describe this as an 'all comer' population, but don't define what this means. The ERG infers that it is likely to mean a population representative of community practice. Choueiri et al[66] state that the cooperative group setting (which they imply is relevant to the CABOSUN trial) is more akin to community practice. The ERG considers that this is a plausible explanation for differences between the sunitinib PFS results of the trial compared to other trials.

Another finding from the CABOSUN trial was that there was a statistically significant difference in the ORR between cabozantinib and sunitinib, favouring cabozantinib. All responses were classed as a 'confirmed partial response', and there were no confirmed complete responders in either study group. Expert clinical advice to the ERG suggests that a complete response would not necessarily be expected in an intermediate or poor risk patient group, and that genuine complete responders to these agents would be relatively unusual.

7.2 Summary of cost effectiveness issues

In the company's analysis the direct comparison of cabozantinib with sunitinib, based on extrapolation of OS, PFS and TTD curves from CABOSUN, gave an ICER of £37,793 per QALY gained. The indirect comparison, with OS and PFS extrapolations based on the fractional polynomial ITC, gave an ICER of £31,538 for cabozantinib compared with sunitinib and £48,451 for cabozantinib compared with pazopanib. In this analysis, pazopanib had lower a lower mean cost and higher mean QALYs than sunitinib: sunitinib is dominated. The company's analysis of uncertainty identifies the OS curves and the cost of cabozantinib as the main drivers of cost-effectiveness.

The ERG identified and corrected some errors and inconsistencies in the company's submitted model the most significant of which was a coding error in QALY calculations that had the effect of underestimating QALYs for each treatment. This resulted in lower ICERs for the company's base cases: £31,956 per QALY for the direct comparison of cabozantinib with sunitinib; and for the ITC analysis, £40,757 for cabozantinib compared with pazopanib and £26,182 compared with sunitinib.

The ERG identified a number of uncertainties in the company's model and identified an alternative set of assumptions and input parameters relating to the method of fitting the OS curves, the time horizon and duration of effects, and health state utilities, adverse effects and costs.

The ERG-preferred analyses gave higher ICER estimates: £65,742 for cabozantinib compared with pazopanib and £41,465 compared with sunitinib. As in the company base case, we estimate that sunitinib is dominated by pazopanib due to its higher cost and similar effectiveness. However, this result was sensitive to some cost and resource use assumptions. By assumption, our preferred analysis gave the same life expectancy with sunitinib as with pazopanib, yielding very similar QALY estimates. Cabozantinib has a modest survival advantage and a larger effect on progression free survival and hence QALYs. We believe that these results appropriately reflect evidence from CABOSUN and COMPARZ. The results were generally robust, with the ICERs remaining above £30,000 per QALY gained for all of the scenarios that we tested.

The above analyses include existing PAS discounts for cabozantinib, sunitinib and pazopanib for first-line treatments. However, they exclude these arrangements and other existing PAS discounts for subsequent treatment after failure of first line treatment. We present results for the ERG-corrected company base case and scenarios and for ERG additional analysis in a confidential addendum to this report.

8 REFERENCES

1. Escudier B, Porta C, Schmidinger M, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2016;27(suppl_5):v58-v68. doi: 10.1093/annonc/mdw328
2. Cancer Research UK. Kidney cancer [updated 14/02/2018.
3. Kim DY, Wood CG, Karam JA. Treating the two extremes in renal cell carcinoma: management of small renal masses and cytoreductive nephrectomy in metastatic disease. *American Society of Clinical Oncology educational book American Society of Clinical Oncology Meeting* 2014:e214-21. doi: 10.14694/EdBook_AM.2014.34.e214 [published Online First: 2014/05/27]
4. Ljungberg B, Hanbury DC, Kuczyk MA, et al. Renal Cell Carcinoma Guideline. *European urology* 2007;51(6):1502-10. doi: 10.1016/j.eururo.2007.03.035
5. National Cancer Registration and Analysis Service. TNM stage group by CCG by tumor type for 10+3 tumour types, 2015 [Available from: http://www.ncin.org.uk/publications/survival_by_stage accessed 21/02/2018.
6. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Kidney Cancer, version 2.2018. *https://wwwnccnorg/professionals/physician_gls/defaultaspx* 2017
7. Heng DY, Xie W, Regan MM, et al. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. *The Lancet Oncology* 2013;14(2):141-8. doi: 10.1016/s1470-2045(12)70559-4 [published Online First: 2013/01/15]
8. Motzer RJ, Bacik J, Schwartz LH, et al. Prognostic factors for survival in previously treated patients with metastatic renal cell carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2004;22(3):454-63. doi: 10.1200/jco.2004.06.132 [published Online First: 2004/01/31]

9. Harding G, Cella D, Robinson D, Jr., et al. Symptom burden among patients with renal cell carcinoma (RCC): content for a symptom index. *Health and quality of life outcomes* 2007;5:34. doi: 10.1186/1477-7525-5-34 [published Online First: 2007/06/16]
10. Cella D. Beyond traditional outcomes: improving quality of life in patients with renal cell carcinoma. *The oncologist* 2011;16 Suppl 2:23-31. doi: 10.1634/theoncologist.2011-S2-23 [published Online First: 2011/03/05]
11. Gupta K, Miller JD, Li JZ, et al. Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. *Cancer treatment reviews* 2008;34(3):193-205. doi: 10.1016/j.ctrv.2007.12.001 [published Online First: 2008/03/04]
12. National Institute for Health and Care Excellence. Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma. Technology appraisal guidance (TA169). London, 2009.
13. National Institute for Health and Care Excellence. Pazopanib for the first-line treatment of advanced renal cell carcinoma. Technology appraisal guidance (TA215). London, 2013.
14. National Institute for Health and Care Excellence. Cabozantinib for previously treated advanced renal cell carcinoma. Technology appraisal guidance (TA463). London: NICE, 2017.
15. Swinburn P, Lloyd A, Nathan P, et al. Elicitation of health state utilities in metastatic renal cell carcinoma. *Current medical research and opinion* 2010;26(5):1091-6. doi: 10.1185/03007991003712258 [published Online First: 2010/03/17]
16. Office for National Statistics. Cancer registration statistics, England: first release, 2016 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2016 accessed 21/02/2018.
17. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. *Nature reviews Urology* 2010;7(5):245-57. doi: 10.1038/nrurol.2010.46 [published Online First: 2010/05/08]
18. Verine J, Pluvinage A, Bousquet G, et al. Hereditary renal cancer syndromes: an update of a systematic review. *European urology* 2010;58(5):701-10. doi: 10.1016/j.eururo.2010.08.031 [published Online First: 2010/09/08]
19. Koneru R, Hotte SJ. Role of cytokine therapy for renal cell carcinoma in the era of targeted agents. *Current oncology (Toronto, Ont)* 2009;16 Suppl 1:S40-4. [published Online First: 2009/05/30]

20. National Institute for Health and Care Excellence. Bevacizumab (first-line), sorafenib (first- and second-line), sunitinib (second-line) and temsirolimus (first-line) for the treatment of advanced and/or metastatic renal cell carcinoma. Technology appraisal guidance (TA178). London, 2009.
21. National Institute for Health and Care Excellence. Renal cancer overview. https://pathwaysniceorguk/pathways/renal-cancer.
22. National Institute for Health and Care Excellence. Tivozanib for treating advanced renal cell carcinoma (TA512). London, 2018.
23. Powles T, Albiges L, Staehler M, et al. Updated European Association of Urology Guidelines Recommendations for the Treatment of First-line Metastatic Clear Cell Renal Cancer. *European urology* 2017 doi: 10.1016/j.eururo.2017.11.016 [published Online First: 2017/12/11]
24. Choueiri TK, Halabi S, Sanford BL, et al. Cabozantinib Versus Sunitinib As Initial Targeted Therapy for Patients With Metastatic Renal Cell Carcinoma of Poor or Intermediate Risk: The Alliance A031203 CABOSUN Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2017;35(6):591-97. doi: 10.1200/jco.2016.70.7398 [published Online First: 2017/02/16]
25. Choueiri TK, Hessel C, Halabi S, et al. Progression-free survival (PFS) by independent review and updated overall survival (OS) results from Alliance A031203 trial (CABOSUN): Cabozantinib versus sunitinib as initial targeted therapy for patients (pts) with metastatic renal cell carcinoma (mRCC). *Annals of Oncology* 2017;28:1.
26. Cabozantinib versus sunitinib for previously untreated patients with advanced renal cell carcinoma (RCC) of intermediate or poor risk: Subgroup analysis of progression-free survival (PFS) and objective response rate (ORR) in the Alliance A031203 CABOSUN trial. ASCO GenitoUrinary Cancer Symposium 2018 February 10, 2018; San Francisco, CA.
27. Exelixis Inc. Clinical Study Report: A031203 Randomized Phase II Study Comparing Cabozantinib (NSC #761968 and IND #116059) with Commercially Supplied Sunitinib in Subjects with Previously Untreated Locally Advanced or Metastatic Renal Cell Carcinoma. San Francisco, 2017.
28. Choueiri TK, Halabi S, Sanford B, et al. CABOzantinib versus SUNitinib (CABOSUN) as initial targeted therapy for patients with metastatic renal cell carcinoma (mRCC) of poor and intermediate risk groups: Results from ALLIANCE A031203 trial. *Annals of Oncology* 2016;27(suppl_6):LBA30_PR-LBA30_PR. doi: 10.1093/annonc/mdw435.23

29. Chopra M. Annual Congress of the European Society for Medical Oncology (ESMO): Copenhagen, Denmark; 7–11 October 2016, 2016.
30. Choueiri TK, Hessel C, Halbi S, et al. Progression-free survival by independent review and overall survival update for the Alliance A031203 CABOSUN trial of cabozantinib vs sunitinib in metastatic renal cell carcinoma. European Society for Medical Oncology Congress. Madrid, Spain, 2017.
31. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York Publishing Services Ltd., 2009.
32. Pond GR. Statistical issues in the use of dynamic allocation methods for balancing baseline covariates. *British Journal of Cancer* 2011;104(11):1711-15. doi: 10.1038/bjc.2011.157
33. Wilson MK, Collyar D, Chingos DT, et al. Outcomes and endpoints in cancer trials: bridging the divide. *The Lancet Oncology* 2015;16(1):e43-e52. doi: 10.1016/S1470-2045(14)70380-8
34. Motzer RJ, Hutson TE, Cella D, et al. Pazopanib versus Sunitinib in Metastatic Renal-Cell Carcinoma. *New England Journal of Medicine* 2013;369(8):722-31. doi: 10.1056/NEJMoa1303989
35. Motzer RJ, Hutson TE, McCann L, et al. Overall Survival in Renal-Cell Carcinoma with Pazopanib versus Sunitinib. *New England Journal of Medicine* 2014;370(18):1769-70. doi: 10.1056/NEJMc1400731
36. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Research synthesis methods* 2010;1(3-4):258-71. doi: 10.1002/jrsm.25 [published Online First: 2010/07/01]
37. ECOG-ACRIN Cancer Research Group. ECOG Performance Status 2018 [Available from: http://ecog-acrin.org/resources/ecog-performance-status accessed 13/3/18.
38. GlaxoSmithKline. Study VEG108844, a study of pazopanib versus sunitinib in the treatment of subjects with locally advanced and/or metastatic renal cell carcinoma.: GlaxoSmithKline, 2013.
39. GlaxoSmithKline. Study VEG108844, a study of pazopanib versus sunitinib in the treatment of subjects with locally advanced and/or metastatic renal cell carcinoma: overall survival update. Report dated 13 Mar 2014.: GlaxoSmithKline, 2014.
40. Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Medical Research Methodology* 2011;11(1):61. doi: 10.1186/1471-2288-11-61

41. National Institute for Health and Care Excellence. Fulvestrant for the treatment of locally advanced or metastatic breast cancer. Technolgy appraisal guidance (TA239). London: NICE, 2011.
42. National Institute for Health and Care Excellence. Fulvestrant for untreated locally advanced or metastatic oestrogen-receptor positive breast cancer. Technolgy appraisal guidance (TA503). London: NICE, 2018.
43. MRC Biostatistics Unit. DIC: Deviance Information Criteria 2017 [Available from: https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/ accessed 12/3/18.
44. Gelman A, Shirley K. Chapter 6: Inference from simulations and monitoring convergence. In: Brooks S, Gelman A, Jones G, et al., eds. Handbook of Markov Chain Monte Carlo. Boca Raton, Florida: Chapman and Hall 2001.
45. Ara R, Brazier J. Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. *Value In Health* 2010;13(5) doi: https://doi.org/10.1111/j.1524-4733.2010.00700.x
46. Latimer N. NICE DSU technical support document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. NICE Decision Support Unit. 2013. *NICE DSU technical support document* 2013
47. Ruiz-Morales JM. First-line sunitinib versus pazopanib in metastatic renal cell carcinoma: Results from the International Metastatic Renal Cell Carcinoma Database Consortium. *European Journal of Cancer* 2016;65:102-08. doi: http://dx.doi.org/10.1016/j.ejca.2016.06.016
48. National Institute for Health and Care Excellence. Single technology appraisal (STA). Pazopanib (Votrient®) for the first-line treatment of patients with advanced renal cell carcinoma (RCC). Manufacturer submission. 16 April 2010 04/16/2010. https://www.nice.org.uk/guidance/ta215/documents/renal-cell-carcinoma-first-line-metastatic-pazopanib-manufacturer-submission-submission2 (accessed January 2018).
49. National Institute for Health and Care Excellence. TA169. Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma 25 March 2009 [Available from: https://www.nice.org.uk/guidance/ta169 accessed January 2018.
50. Amdahl J. Cost-effectiveness of pazopanib compared with sunitinib in metastatic renal cell carcinoma in Canada. *Current Oncology* 2016;23(4):e340-e54.
51. Choueiri TK, Escudier B, Powles T, et al. Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2015;373(19):1814-23. doi: 10.1056/NEJMoa1510016 [published Online First: 2015/09/26]

52. Curtis L, Burns A. Unit Costs of Health and Social Care 2017. 2017 ed. Canterbury, Kent: Personal Social Services Research Unit. University of Kent, 2017.
53. NHS Improvement. NHS Reference Costs 2016/2017. 2017 [Available from: https://improvement.nhs.uk/resources/reference-costs/ accessed 10/4/2018.
54. National Institute for Health and Care Excellence. Everolimus for advanced renal cell carcinoma after previous treatment. Technology appraisal guidance (TA432). London: NICE, 2017.
55. National Institute for Health and Care Excellence. Lenvatinib with everolimus for previously treated advanced renal cell carcinoma Technology appraisal guidance (TA498). London: NICE, 2018.
56. National Institute for Health and Care Excellence. Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment. Technology appraisal guidance (TA333). London: NICE, 2015.
57. National Institute for Health and Care Excellence. Nivolumab for previously treated advanced renal cell carcinoma. Technology appraisal guidance (TA417). London: NICE, 2016.
58. Hudes G, Carducci M, Tomczak P, et al. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *New England Journal of Medicine* 2007;356(22):2271-81. doi: DOI: 10.1056/NEJMoa066838
59. Choueiri TK, Escudier B, Powles T, et al. Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *The Lancet Oncology* 2016;17(7):917-27. doi: https://doi.org/10.1016/S1470-2045(16)30107-3
60. Rini BI, Halabi S, Rosenberg JE, et al. Bevacizumab plus interferon alfa compared with interferon alfa monotherapy in patients with metastatic renal cell carcinoma: CALGB 90206. *Clinical Oncology* 2008;26(33):5422. doi: https://dx.doi.org/10.1200%2FJCO.2008.16.9847
61. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. London: Nuffield Trust, 2014.
62. Inflation.EU. Average inflation rate in Great Britain [Available from: http://www.inflation.eu/inflation-rates/great-britain/historic-inflation/cpi-inflation-great-britain.aspx. accessed 9/8/16.
63. Rini. BI, Vogelzang. NJ. Future Challenges for Drug Development in Renal Cell Carcinoma. *Journal of Clinical Oncology* 2017;35(6):577-79. doi: 10.1200/jco.2016.71.0673

64. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus Interferon Alfa in Metastatic Renal-Cell Carcinoma. *New England Journal of Medicine* 2007;356(2):115-24. doi: 10.1056/NEJMoa065044

65. Ko JJ, Choueiri TK, Rini BI, et al. First-, second-, third-line therapy for mRCC: benchmarks for trial design from the IMDC. *British Journal Of Cancer* 2014;110:1917. doi: 10.1038/bjc.2014.25

66. Choueiri TK, Halabi S, Morris MJ, et al. Reply to B. Rini et al and S. Buti et al. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2017;35(16):1859-60. doi: 10.1200/jco.2017.72.2629 [published Online First: 2017/05/27]

67. Schnadig ID, Hutson TE, Chung H, et al. Dosing patterns, toxicity, and outcomes in patients treated with first-line sunitinib for advanced renal cell carcinoma in community-based practices. *Clinical genitourinary cancer* 2014;12(6):413-21. doi: 10.1016/j.clgc.2014.06.015 [published Online First: 2014/08/29]

9 APPENDICES

9.1 ERG critical appraisal of the ITC

Criterion	ERG assessment
ITC purpose	
1. Are the ITC results used to support the evidence for the clinical effectiveness of the intervention?	Yes, for the comparison of cabozantinib with pazopanib.
2. Are the ITC results used to support the evidence for the cost-effectiveness of the intervention?	Yes. OS and PFS results from the ITC are used directly in the economic model to inform the estimates of cost effectiveness.
Evidence selection	
3. Are inclusion/exclusion criteria adequately reported?	Yes. CS Table 4 lists the inclusion criteria. These criteria include a broader list of treatments than in the NICE scope. The CS notes that the systematic review was conducted from a global perspective and consequently included additional comparator treatments not specified in the NICE scope. Subsequent restriction to only comparator treatments in the scope resulted in inclusion of 2 studies (n=9 records), the CABOSUN trial and the COMPARZ trial.
4. Is quality of the included studies assessed?	Yes, for the 2 studies in the restricted ITC network (Table 15 in Appendix D1.3, and Figure 41 and 42 in Appendix D1.1), using the standard criteria recommended by NICE.
Methods – statistical model	
5. Is the statistical model described?	Yes. Three types of statistical method are used: (1) Indirect comparison of parametric survival curves using methodology developed by Ouwens et al (2010) (2) Parametric models with fractional polynomial distributions using methodology developed by Jansen et al (2011). (3) A network meta-analysis supplementary method comparing hazard ratios using a fixed effects model, for intermediate risk and poor risk subgroups. Methods 1 and 2 are used to inform the economic model.
6. Has the choice of outcome measure used in the analysis been justified?	Yes, OS and PFS are key outcomes in cancer survival modelling.
7. Has a structure of the network been provided?	Diagrams illustrating the networks are provided in the CS: Figure 9 shows the primary evidence network for potential meta analysis (i.e. based on the broader inclusion criteria). CS Figure 11 shows the restricted evidence network containing the 2 included RCTs. CS Appendix D1.1 shows the networks used in the NMA supplementary method of HRs (CS Figures 43 to 48).
8. Is homogeneity considered?	Yes, discussed in CS section B.2.9 and Appendix D1.1. A feasibility assessment is described to assess differences in study and patient characteristics within and between treatment comparisons. CS Table 22 tabulates risk category and performance status details between the 2 included trials.

9. Are the studies homogenous in terms of patient characteristics and study design?	No. The CABOSUN trial included only patients at intermediate or poor risk, whilst the COMPARZ study included patients with favourable, intermediate and poor risk classifications. The distribution of patients between risk classifications is different between the two trials. The CS acknowledges that the differences in distribution of risk category is the variable that most affects survival. There were slight differences between trials in the number of metastatic sites detected (≥3 sites: 32% to 41% by treatment arm in CABOSUN; 42% to 44% by treatment arm in COMPARZ). (CS Appendix Table 11). Just over a third of patients in CABOSUN had bone metastases at baseline (36%-37% by trial arm) compared to 15%-20% (by trial arm) of patients in COMPARZ. CABOSUN was a small phase II RCT (n=157 patients), whilst COMPARZ was a larger phase III RCT (n=1110 patients randomised).
10. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. sub group analysis, sensitivity analysis, meta-regression)	The network meta-analysis supplementary method presented in Appendix D conducted separate NMAs in intermediate risk and poor risk subgroups, including comparators outside of the scope. This NMA is reported to be additional to the Ouwens et al and fractional polynomial analyses, specifically to explore the results of subgroup analyses compared to the overall study populations. However, unlike the other two analyses, this method does assume proportional hazards.
11. Is the assumption of similarity stated?	Yes – the CS discusses similarity ('assessment of heterogeneity' in Section B.2.9), describing the similarities and differences between the two trials, but does not explicitly state whether the assumption of similarity holds. In Appendix D1.1 it is stated that the "populations in CABOSUN and COMPARZ are different".
12. Is any of the programming code used in the statistical programme provided (for potential verification)?	Yes, provided in CS Appendix D1.1.
Sensitivity analysis	
13. Does the study report sensitivity analyses?	No.
Results	
14. Are the results of the ITC presented?	Yes. Most of the detail is in Appendix D1.1, with presentation of a series of graphs showing fitted survival curves for the Ouwens et al model and the fractional polynomials models, for both OS and PFS and for random and fixed effects models, where conducted (Figures 1 to 40). However, the CS did not report hazard plots depicting the time-varying hazard ratios and their credible intervals from the fractional polynomials models. The ERG requested these plots from the fractional polynomials analysis from the company (clarification question A22 and A23).

15. Does the study describe an assessment of the model fit?	Yes. Fit statistics for the Ouwens et al and the fractional polynomial methods, for OS and PFS, are presented in CS section B.2.9 (Tables 23 and 24). The Deviance Information Criteria (DIC) was used to select the model with the best fit, with a lower posterior mean DIC indicating a better fit. This is a standard approach to assessing Bayesian model fit. The CS does not report any other considerations in relation to model fit (e.g. plausibility of modelled distribution).
16. Has there been any discussion around the model uncertainty?	No. The ERG requested the company to provide the credible intervals for the time-varying hazard ratios estimated by the fractional polynomial model (clarification question A22) to assess the degree of uncertainty. The Ouwens et al models were conducted using fixed effect and random effects, and the fractional polynomials models were conducted using only fixed effects. The company were requested to supply the random effects fractional polynomial model (clarification question A23).
17. Are the point estimates of the relative treatment effects accompanied by some measure of variance such as confidence intervals?	No. However, the Ouwens et al and fractional polynomial methods do not estimate a single point estimate, such as a constant hazard ratio. For example, the fractional polynomial method estimates time-varying hazards over time. The ERG requested the company to provide hazard ratios and credible intervals for each interval of the follow-up time period for the fractional polynomial models (clarification question A22).
Discussion - overall results	
18. Does the study discuss both conceptual and statistical heterogeneity?	Yes. Conceptual (clinical) heterogeneity is discussed (see above). Statistical heterogeneity was not relevant as the ITC in the restricted network included only two trials, linked together by a common comparator arm.
Discussion - validity	
19. Are the results from the indirect/NMA compared, where possible, to those just using direct evidence?	No. This was not necessary as there are no comparisons informed by both direct and indirect evidence.

9.2 Critical appraisal of the COMPARZ trial

The table below presents the company's and the ERG's critical appraisal of the COMPARZ trial.[34]

NICE QA Criteria for RCT	CS response	ERG response
1. Was the method used to generate random allocations adequate?	Yes	Unclear
The trial publication states that patients were randomly assigned to one of the two study drugs in a 1:1 ratio in permuted blocks of four. The CSR states that the randomisation schedule was generated by GSK Statistics and Programming Department (page 45). However, it does not state the exact method used to generate the schedule.		
2. Was the allocation adequately concealed?	Not clear	Yes

Comments: The CSR states that an interactive voice response system was used (section 5.3). All patients were entered into this system after baseline assessment and the randomisation schedule was then generated centrally. It appears that study sites called the interactive voice system to request randomisation when required. Thus sites could not have known in advance the next random allocation in the sequence.		
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Yes	Yes
Comments: There do not appear to be any notable differences between the groups in demographic or clinical characteristics (Supplementary Table S3 to the trial journal publication ([34]).		
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	No	No
Comments: The trial was open-label. However, imaging data were re-evaluated by an independent review committee whose members were unaware of the treatment assignments to assess the primary end point and tumour response.		
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No	No
Comments: The number of treatment discontinuations was similar between the two groups, and the reasons for discontinuations were broadly similar (Supplementary Figure S2 to the trial journal publication[34])		
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No
Comments: The ERG checked the objectives (outcomes) stated in the CSR and outcome data are reported for each of them.		
7. Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes Yes Unclear
Comments: Efficacy data were analysed in the intention-to-treat population (all patients who underwent randomisation). However, the ERG notes that for patient reported outcomes (HRQoL and symptoms) the number of patients analysed is lower than the number randomised. It is not clear how missing data were handled (see Table 2 in the trial journal publication.[34]		

9.3 Description and critique of ITC method 3: Network meta-analysis supplementary method

CS Appendix D1.1 reports brief details of what the CS describes as a supplementary NMA of cabozantinib compared to sunitinib, pazopanib, interferon-alfa, sorafenib, bevacizumab in combination with interferon-alfa, temsirolimus, and tivozanib. The CS reports that non-scope treatments were included in this network because the analysis was conducted for a non-English perspective.

The CS states that Kaplan-Meier data results were not available for intermediate and poor prognostic risk groups separately from the ITT population (CS Section B.2.9, page 56). To further explore the impact of differences in the subgroup data, an additional analysis was carried

out on hazard ratios (HRs). Unlike Kaplan-Meier data, HRs were available by subgroup and these were compared despite the violation of the proportional hazards assumption (see Appendix D for further details).

Separate evidence networks were constructed based on RCC risk groups: intermediate risk, poor risk, and the overall population of patients. The networks include comparators inside and outside of the NICE appraisal scope, and they vary in size according to population group (e.g. overall population, risk subgroup) and outcome measure. The CS describes the set of studies as heterogeneous in terms of RCC risk groups, with some studies including patients with favourable RCC risk. The CS does not provide any further details of the characteristics of the included studies, but does tabulate the OS and PFS HRs for the ITT populations, and intermediate and poor risk subgroups for each of the trials (CS Appendix D1.1 Table 14). The results of the NMA are presented as a series of fixed effect forest plots showing the HRs for each of the treatments compared to cabozantinib (CS Appendix D1.1 Figure 49 to Figure 51).

The ERG urges caution in the interpretation of these analysis as they assume that the proportional hazards assumption holds in the CABOSUN trial (and other trials in the network), yet as discussed above, this assumption is not supported by the OS curves in the trial. Furthermore, no assessment of heterogeneity or consistency has been provided for the trials in the networks, and the networks use data from subgroups of the randomised patient populations (the size of which are unspecified in the CS), therefore can be considered observational evidence, and likely underpowered due to small sample sizes. Furthermore, very little information is given on the statistical methods used to conduct this analysis.

9.4 Additional results of the ITC

Fixed effect ITC fractional polynomial model – additional results

In section 3.3.7 of this report we reported the results of the best fitting fractional polynomial models. Here we summarise the results of the other fractional polynomial models tested. There were some differences in results between the different fractional polynomial models:

Progression free survival

- **First order model results**. In three of the models (P=0; P=0.5; P=1) there was a slight decline in the time-varying HR curves over time, from around 0.5 to around 0.3-0.4 (NB.

The ERG was unable to cross-check the HR plots with the tabulated HRs for first order PFS pazopanib as Table 5 appears to be a duplicate of Table 3, which is the tabulated HRs for first order OS pazopanib (clarification question A22)). Credible intervals tended to increase markedly over time and exceeded 1. In the other two models (P=-1; P=-0.5), the HR curves were flat at around 0.5 for the entire follow-up period, indicating little change in HRs over time.

- **Second order model results**. In all models the time-varying HRs for both comparisons increase sharply from zero within the first three months to reach a plateau of around 0.5, then decline slightly over time to around 0.4. Credible intervals tended to increase markedly over time in all models and exceeded 1, though the intervals in the best fitting model (p1=-1, p2=-1) are less wide than the other models. The results of the best fitting fractional polynomial as used in the economic model are therefore consistent with the other second order models, though with less uncertainty.

Overall survival

- **First order model results**. In most of the first order fractional polynomial models the time-varying HRs curves are reasonably straight over time (at around 0.7-0.8), indicating a constant HR. The exception is first order fractional polynomial model p=1 in which the curves decline slightly over time from around 0.8-1.0 to around 0.6. Fractional polynomial first order model (p=-0.5) appears to be an outlier as pazopanib has a slightly higher HR compared to sunitinib by an order of approximately 0.1 (around 0.6 compared with around 0.5, respectively) and the credible intervals are wider than all the other first order models.
- **Second order model results.** The second order fractional polynomial model curves have a distinctly different shape to the first order curves. As was the case for PFS, the time-varying HRs for both comparisons increase sharply from zero within the first six months to reach a plateau, then decline slightly over time. The exception is the best fitting fractional polynomial model (p1=-1, p2=-1) where the HRs remain generally constant (and higher than the other models) once they have peaked.

Random effects fractional polynomial model results

The fractional polynomial ITC results presented in the CS were based on a fixed effect model. For comparison, the company were asked to provide fractional polynomial results based on a random effects model (clarification question A23). The ERG crosschecked the results of the fixed effect and random effects fitted fractional polynomial curves (for the restricted network

only). In all but one of the fractional polynomial models, the results appeared similar between the fixed effect and the random effects models. The exception was the OS 1st order (p=-0.5) model where the random effects model (Figure 39, clarification question A23) had higher curves for all three treatments compared to the fixed effect model (CS Appendix D1.1 Figure 24). It is not clear why this is the case. Importantly, the fitted curves for random effects and fixed effect models in the best-fitting fractional polynomial model (used to inform the economic model) were similar to each other, indicating that the inclusion of additional evidence did not change the results.

ITC results for the wider evidence network

The company were asked to provide ITC results based on the wider network of 13 RCTs that included studies of additional treatments not within the scope of the appraisal (clarification question A26) (see section 3.1.7.1 for a discussion of this network). The aim was to check whether the results for the comparison between cabozantinib, sunitinib and pazopanib were different when a wider network containing other treatment comparisons was used.

The company point out that there is considerable clinical heterogeneity in this network, citing the TARGET study of sorafenib versus placebo as comprising a mostly pre-treated population. They also mention that there were differences in the extent of patient crossover in some trials. The company has not presented tabulated characteristics of these studies to allow an assessment of clinical heterogeneity, but the ERG agrees that it is plausible that clinical heterogeneity would exist in this wider set of studies.

The ERG cross-checked the results of the wider and the restricted networks for the Ouwens et al ITC fixed effect and random effects models. The results were similar in all cases except for the exponential model where there were bigger differences in the fitted survival curves between pazopanib and sunitinib (whereas in the restricted network they were similar). The reason for this disparity between the networks is not clear. Results from Gompertz survival models in the wider network were not supplied in response to clarification question A26 so the ERG are unable to check the consistency of results for this model between the networks.

The ERG cross-checked the results of the wider and the restricted networks for the fixed effects fractional polynomials models. In all but one of the models, the results appeared similar between the wider and restricted networks. The shape of the fitted PFS survival curves for

cabozantinib, sunitinib and pazopanib in first order model (p=-0.5) of the restricted network (CS Appendix D1.1 Figure 34) did not correspond to the corresponding curves in the wider network (Figure 131, clarification question A26). It is not clear why this is the case. Also, one of the fractional polynomial second order models (P1=-0.5, P2=0) based on the wider network did not converge, so it is not possible to compare its results with the corresponding model in the restricted network. Importantly, the fitted curves for the wider and the restricted networks in the best-fitting fractional polynomial model (used to inform the economic model) were similar to each other, indicating that the inclusion of additional evidence did not change the results.