CONFIDENTIAL UNTIL PUBLISHED

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

(Erratum – errors corrected following factual accuracy check of ERG report)

Pembrolizumab in combination with axitinib for untreated advanced renal cell carcinoma

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors
Dr Keith Cooper, Senior Research Fellow, SHTAC
Dr Emma Loveman, Senior Researcher, Effective Evidence LLP
Mr Olu Onyimadu, Research Fellow, SHTAC
Mr David A. Scott, Director, Diligent Agile Synthesis Limited
Dr Jill Colquitt, Senior Researcher, Effective Evidence LLP
Dr Jonathan Shepherd, Principal Research Fellow, SHTAC

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 25th September 2019

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 127748

Declared competing interests of the authors

None

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 also thank Dr Geoff Frampton (SHTAC) for acting as an internal editor for the draft report.

Declared competing interests of the authors and clinical advisors

The authors declare no competing interests. Professor Robert Hawkins declares receiving speaker fees from the following companies (treatments): EUSA Pharma (tivozaninb); Ipsen (cabozantinib); Novartis (pazopanib); Pfizer (sunitinib and axitinib). Dr Matthew Wheater declares receiving support from Merck Sharpe & Dohme for conference attendance in June 2019, in which he participated in a general discussion with the company about treatment for renal cell carcinoma (not specifically related to the company's evidence submission to NICE on pembrolizumab plus axitinib). Dr Wheater also declares receiving speaker fees from Pfizer.

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:

Cooper, K., Loveman, E., Onyimadu, O., Scott, D., Colquitt, J., Shepherd, J., Pembrolizumab in combination with axitinib for untreated advanced renal cell carcinoma: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2019.

Contributions of authors

Keith Cooper critically appraised the economic evaluation, and drafted the report. Emma Loveman critically appraised the clinical effectiveness review and drafted the report. Olu Onyimadu critically appraised the economic evaluation, and drafted the report. David Scott critically appraised the indirect treatment comparison clinical and drafted the report. Jill Colquitt critically appraised the clinical effectiveness review and drafted the report. Jonathan Shepherd critically appraised the clinical effectiveness review, drafted the report, project managed the report and is the project guarantor.

Word count: 40,593

Commercial in confidence (CIC) information in blue
Academic in confidence (AIC) information in yellow.

LIST OF ABBREVIATIONS

AE	Adverse event
AIC	Academic in confidence
AJCC	American Joint Committee on Cancer
BICR	Blinded independent central review
CIC	Commercial in confidence
CPS	Combined Positive Score
CR	Complete Response
CS	Company submission
CSR	Clinical study report
CT	Computerised tomography
DCR	Disease Control Rate
DIC	Deviance information criteria
DOR	Duration of Response
ECOG	Eastern Cooperative Oncology Group
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 items
EPAR	European Public Assessment Report
ERG	Evidence Review Group
FDA	Food and Drug Administration
FP	Fractional polynomial
HR	Hazard ratio
HRQoL	Health related quality of life
IFN-α	Interferon alpha
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
KPS	Karnofsky Performance Status
MCMC	Memorial Sloan-Kettering Cancer Center
N/A	Not applicable
NCCN	National Comprehensive Cancer Network
NE	Not estimable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
OS	Overall survival
ORR	Objective response rate
PAS	Patient access scheme
PD	Progressed disease
PD-L1	Programmed death-ligand 1
PF	Progression free

PFS	Progression free survival
PH	Proportional hazards
PR	Partial Response
QALY	Quality adjusted life year
QoL	Quality of life
RCC	Renal cell carcinoma
RECIST	Response evaluation criteria in solid tumours
RTKs	Receptor tyrosine kinases
SAE	Serious adverse event
SD	Stable Disease
SLR	Systematic literature review
SmPC	Summary of Product Characteristics
TEAE	Treatment emergent adverse events
TOT	Time on treatment
TKI	Tyrosine kinase inhibitor
TNM	Tumour Node Metastasis
TTD	Time to treatment discontinuation
VAS	Visual analogue scale
VEGF	Vascular endothelial growth factor (VEGF)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Scope of the company submission

The company's decision problem is as follows:

- Population: Adults with untreated advanced RCC
- Intervention: Pembrolizumab in combination with axitinib
- Comparators: Tivozanib; pazopanib; sunitinib; cabozantinib (for disease that is defined as intermediate or poor risk)
- Outcomes: overall survival (OS); progression-free survival (PFS); objective response rate (ORR); adverse events (AE); health-related quality-of-life (HRQoL).

The company's decision problem is largely consistent with the NICE scope.

Summary of submitted clinical effectiveness evidence

The company conducted a broad literature review to meet the needs of multiple countries. Studies providing direct and indirect evidence relevant to the target population for the current scope were selected during the final stages of the review.

One randomised controlled trial (RCT) of pembrolizumab plus axitinib versus one of the scoped comparators, sunitinib, was identified (KEYNOTE-426). Pembrolizumab 200mg was administered every three weeks by IV infusion for up to 35 doses (about 24 months) and axitinib 5 mg was administered twice daily orally. Sunitinib 50mg was administered daily orally, four weeks on, two weeks off. Treatments were continued until progressive disease was confirmed or unacceptable adverse events. A total of 861 participants with previously untreated locally advanced/metastatic clear cell renal cell carcinoma (RCC) were included. The trial was undertaken in 16 countries and 37% of participants were from Europe. The number randomised in the UK is unclear. The ERG notes there is a risk of performance bias in the trial but risk of bias from other sources is low.

The main results presented in the CS and used in the economic model are from the first planned interim analysis (August 2018 data-cut), after a median follow-up of 13.2 months in the intervention arm and 12.1 months in the comparator arm. Efficacy testing was stopped when this analysis showed statistically significant improvement in both co-primary endpoints [progression-free survival (PFS) and overall survival (OS)] and the key secondary endpoint [objective response rate (ORR)]. Results from a second (unplanned) data-cut in January 2019

for US Food and Drug Administration (FDA) with an additional four months follow-up are presented in appendices. Overall survival follow-up of the trial is ongoing. The ERG notes that early stopping of trials can sometimes results in over-estimation of treatment effect. Based on the number of events, the ERG considers that PFS is unlikely to have been overestimated, but OS at the interim analysis is potentially overestimated and should be interpreted with caution due to data immaturity.

KEYNOTE-426 trial results

- Median OS was not reached in either arm, HR 0.53 (95% CI 0.38 to 0.74), p=0.00005. Results from the January 2019 data cut [redacted].
- Median PFS based on blinded independent central review (BICR) was 15.1 months with pembrolizumab plus axitinib and 11.1 months with sunitinib, HR 0.69 (95% CI 0.57 to 0.84, p=0.00014). Results from the January 2019 data cut [redacted].
- Objective response rate (ORR) was 59.3% in the pembrolizumab plus axitinib arm and 35.7% in the sunitinib arm based on BICR according to RECIST 1.1 criteria, a difference of 23.6% (95% CI 17.2 to 29.9, p<0.0001). [redacted].
- Median duration of response (DOR) based on BICR in people with a complete response or partial response was not reached in the pembrolizumab plus axitinib arm and was 15.2 months in the sunitinib arm. Median DOR based on investigator assessment was 18.0 months (range 1.3+ to 18.2+) and 15.2 months (range 1.2+ to 15.4+), respectively ('+' indicates there was no progressive disease by the time of last disease assessment).
- There were no significant differences between treatments for the EQ-5D-3L index, EQ-5D visual analogue scale (VAS) or most functional and symptom scales of the EORTC-QLQ-30 instrument. The exception was a greater worsening of the EORTC-QLQ-30 diarrhoea symptom scale in the pembrolizumab plus axitinib group.
- Subgroup analyses of OS and PFS were consistent with the effect seen in the overall trial population [redacted]. [redacted].The overall rate of adverse events

(AEs) was similar across both arms of the trial. The rate of serious adverse events (SAEs) was higher in the pembrolizumab plus axitinib group; 40.3% of participants reported SAEs in the pembrolizumab plus axitinib arm compared with 31.3% in the sunitinib arm. For drug-related grade 3 to 5 AEs, pembrolizumab plus axitinib had a higher risk of increased ALT, increased AST and diarrhoea. Sunitinib had a higher risk of fatigue, thrombocytopenia and neutropenia among others. The rates are in line with those of axitinib as monotherapy.

Network meta-analysis results

The CS reports two types of Bayesian approaches for indirect comparison of pembrolizumab plus axitinib with other treatments:

- Network meta-analysis (NMA) assuming constant hazards
- NMA assuming time-varying hazards based on fractional polynomials.

These NMAs were reported for OS and PFS outcomes. The NMA assuming constant hazards appears to be the 'primary' indirect comparison method reported in the CS.

The networks are presented as a base case analysis, which included all patients irrespective of baseline RCC risk status, and subgroup analyses for patients at intermediate/poor RCC risk, and patients at favourable RCC risk. The ERG agrees with the decision to conduct a separate NMA for the intermediate/poor RCC risk group, as inclusion in the CABOSUN trial of cabozantinib was restricted to patients in these risk groups, and cabozantinib is recommended by NICE only for patients at intermediate/poor risk (as defined by the IMDC criteria).

The NMA does not inform the economic model for the base case analysis (all patients irrespective of baseline RCC risk status). The NMA informs the economic model for the subgroup analysis comparing pembrolizumab plus axitinib versus cabozantinib.

The trials (n=6) included in the NMA were generally similar in terms of key patient characteristics, and were assessed by the company and the ERG to be at low risk of bias, with the exception of blinding (trials were open-label).

Overall, the ERG considers the methods and assumptions used to conduct the NMAs to have been appropriately exercised, though with some uncertainties due to relatively small data sets, and potential heterogeneity.

The following results are from the constant hazard NMA

- In the base case NMA, pembrolizumab plus axitinib resulted in [redacted] in the duration of PFS compared to all relevant competing interventions: [redacted]).
- In the intermediate/poor risk subgroup NMA, both cabozantinib and pembrolizumab plus axitinib were associated with [redacted] HRs compared to sunitinib ([redacted] indicating [redacted] PFS.
- In the base case NMA, pembrolizumab plus axitinib was associated with [redacted] in the duration of OS compared to pazopanib ([redacted]) and sunitinib ([redacted]). Tivozanib was omitted from this NMA due to lack of data.
- In the intermediate/poor risk subgroup NMA, pembrolizumab plus axitinib was associated with [redacted] in OS compared to sunitinib [redacted] and compared to cabozantinib [redacted]
- Results from the NMA using the January 2019 KEYNOTE-426 data-cut show [redacted] results to the above.

Summary of submitted cost effectiveness evidence

The CS includes:

- a review of published economic evaluations of comparator therapies to pembrolizumab in treating patients with advanced renal cell carcinoma.
- a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of pembrolizumab in combination with axitinib is compared with sunitinib, tivozanib, pazopanib and cabozantinib for adults with untreated advanced RCC.

The company conducted a systematic search of the literature to identify economic evaluations of comparator treatments to pembrolizumab in untreated advanced RCC. The search identified 10 published cost-effectiveness studies, of which nine were conducted from an English, Welsh or British perspective. None of the studies included pembrolizumab plus axitinib, however the ERG identified a cost-utility study by Chen et al. that compared pembrolizumab plus axitinib to sunitinib in patients with RCC in China.

The company developed a model to evaluate the cost-effectiveness of pembrolizumab plus axitinib as first-line treatment for advanced RCC. The model is a partitioned-survival 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 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 was estimated using a partitioned survival approach, based on PFS, and OS curves. Patients enter the PF state on initiation of first-line treatment but may stop treatment at any time due to adverse effects or when their disease progresses. The proportion of patients on first-line treatment is determined by the time to treatment discontinuation (TTD) curves. Some patients then progress to a subsequent treatment with one of the drugs suitable for second-line treatment. The duration of second-line treatment was taken from the clinical trials for each drug, after which patients are assumed to receive supportive care until death.

The submitted model includes analyses for two patient populations:

- The overall population of KEYNOTE-426;
- Subgroup population of patients with intermediate/poor RCC risk status, as defined by IMDC criteria, in the KEYNOTE-426 population.

The PFS, OS and TTD curves for pembrolizumab plus axitinib, and sunitinib were based upon survival data from the KEYNOTE-426 trial. The CS assumes that sunitinib is clinically equivalent to tivozanib and pazopanib, based on similar assumptions made in previous NICE appraisals for this indication. For the subgroup population at intermediate / poor risk, pembrolizumab plus axitinib is compared to cabozantinib using the company's NMA, as no head-to-head comparison was available.

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

- **Cycle length**: 1 week with half cycle correction implemented.
- **Time horizon**: 40 years in the base case
- **Discounting**: 3.5% per year for costs and QALYs
- **Duration of treatment effects**: based on extrapolation of PFS and OS curves fitted to trial data and based on model fit statistics and clinical expert judgement. The persistence of treatment effect throughout the model time horizon was assumed in the company's base case. Treatment waning after 10 years was tested in a scenario.

- **Adverse events**: includes grade 3 and above all-cause adverse events which occur in at least 5% of patients for all first-line treatments. Adverse events related to subsequent treatments are not explicitly modelled.
- **Utility and QALY calculations**: HRQoL estimates evaluated from the KEYNOTE-426 trial are used in the model. Three approaches where used to estimate HRQoL: estimation of utilities based on progression-free and progressed disease states (with or without differentiation by treatment) and the estimation of utilities based on time to death. These approaches are discussed in detail in section 4.3.6. An age-based utility decrement is also applied.
- **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; monitoring and disease management in PF and PD states; treatment of included TEAEs for first-line treatments; and terminal care costs in the last cycle before death.
- **Uncertainty**: the model incorporates 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.

Parametric survival curves were fitted to PFS, OS and TTD data from the KEYNOTE-426 trial. The survival curves used in the company's base case and scenario analyses are summarised in the table below.

Table 1 Survival curves used in the company's economic analyses

Curve	Treatment	CS Base case	CS scenarios
PFS	Pembrolizumab + axitinib Sunitinib	Exponential	Lognormal for P+A, Exponential for S
OS	Pembrolizumab + axitinib Sunitinib	Log-logistic Exponential	Exponential Time varying hazard ratio
TTD	Pembrolizumab + axitinib Sunitinib	Weibull Exponential Exponential	Weibull for P, Log-normal for A, Exponential for S.

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 were taken from the company's KEYNOTE–426 trial using the time-to-death approach. Adverse event disutilities were estimated according to the EQ-5D values collected in the KEYNOTE-426 trial for pembrolizumab plus axitinib versus sunitinib.

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; monitoring and disease management in PF and PD states; treatment of AEs for first-line treatments; and end of life care.

The results of the economic model are presented below using list prices for all drugs as incremental cost effectiveness ratios (ICERs). (NB. cost-effectiveness results based on patient access scheme discount prices for comparator treatments and treatments used in subsequent treatment lines are presented in a separate confidential appendix to this report).

- For the company base case, an ICER of £59,292 per QALY gained is reported for pembrolizumab plus axitinib versus sunitinib.
- For the intermediate/poor risk subgroup analysis, an ICER of £21,452 per QALY gained is reported for pembrolizumab plus axitinib versus cabozantinib.

The company conducted one-way deterministic sensitivity analyses and concluded that the key drivers of the cost-effectiveness results were changes to the distribution used for extrapolating OS and the discount rates for QALYs. The company's scenario analyses found cost-effectiveness results to be most sensitive to the choice of OS curve used in the model. The company's base case probabilistic sensitivity analysis gave an ICER of £59,726 per QALY gained and estimated a 0.3% chance of pembrolizumab plus axitinib is cost-effective at the £30,000 per QALY threshold compared to sunitinib.

End of life criteria and innovation

- The ERG agrees with the company that pembrolizumab plus axitinib does not meet the first end of life criterion in the overall RCC population ("treatment is indicated in patients with a short life expectancy, normally less than 24 months").
- The ERG disagrees with the company that pembrolizumab plus axitinib meets the first end of life criterion in the poor RCC risk subgroup. We consider cabozantinib to be the

NICE recommended treatment for this group rather than sunitinib as referenced in the CS. The company does not provide an explicit rationale for singling out the poor risk group, as opposed to the intermediate / poor risk subgroup. Estimates of cost-effectiveness are not provided in the CS for the poor risk group.

- The ERG is in agreement with the company that pembrolizumab plus axitinib meets the second end of life criterion ("treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment").
- We are therefore of the opinion that pembrolizumab plus axitinib does not fully meet the NICE criteria for being considered as a life-extending treatment for people with a short life expectancy.

The company considers that the innovative immuno-oncology combination regimen of pembrolizumab plus axitinib represents a "step-change" in the management of RCC as it targets both angiogenesis and immune-checkpoint pathways. The CS states that pembrolizumab should be considered innovative by its potential to make a significant and substantial impact in an area of high unmet need. The ERG clinical advisors agree there does remain an element of unmet need and that the rationale for the treatment combination in RCC is made. However, there are other potential treatments that should be considered in relation to pembrolizumab and axitinib, such as avelumab plus axitinib, currently the subject of a separate NICE technology appraisal.

Commentary on the robustness of submitted evidence

Strengths

- The company conducted a reasonable quality systematic review and it is unlikely that any relevant trials have been omitted.
- The evidence for the clinical effectiveness of pembrolizumab plus axitinib is from a large multinational RCT (KEYNOTE-426) comparing the treatment with one of the NICE scoped comparators, sunitinib. The open-label design of the trial means there is a risk of performance bias, but the risk of bias from other sources is low. Outcomes of the trial are appropriate and relevant to the scope.
- The methods and assumptions of the company's NMA are generally appropriate, although uncertainties remain due to the relatively small datasets in subgroup analyses.

- The comparator trials included in the NMA have a low risk of bias, other than that due to their open-label design.
- The structure of the company's economic model reflects the nature of disease progression and the clinical pathway for people with untreated locally advanced or metastatic RCC. The methods used for the economic evaluation are consistent with NICE methodological guidelines and other technology appraisals for treatment for this population.
- EQ-5D utility values were collected in the KEYNOTE-426 trial. These utility values 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.

Weaknesses and areas of uncertainty

- KEYNOTE-426 restricted inclusion to clear cell RCC. It is not clear whether results are generalisable to all types of RCC. However, this is in line with the pivotal phase III trials of comparator treatments which have been the subject of previous NICE appraisals in this indication.
- The majority of participants in KEYNOTE-426 (63%) were not randomised in Europe. The exact number of UK participants is unclear, but was less than 6% of the total randomised.
- Although typical of a phase III trial, the participants are generally younger and fitter than the general population with adults with untreated locally advanced or metastatic RCC.
- Efficacy testing was stopped early at the first interim analysis. Early stopping can sometimes result in over-estimation of treatment effect, in this case it is unlikely that PFS has been over-estimated, but OS results should be viewed with caution.
- There is significant uncertainty over the extrapolation of OS due to OS immaturity of survival data from the KEYNOTE-426 trial. ICERs are sensitive to this uncertainty. We consider that the Weibull distribution is more plausible for the extrapolation of OS and has more conservative survival predictions. We do not agree with the company's use of different survival curves for the extrapolation of the pembrolizumab plus axitinib and sunitinib arms.

Summary of additional work undertaken by the ERG

The ERG's preferred set of assumptions included the following key differences from the company base case:

- **Method of fitting OS curves**. The ERG considers that the Weibull distribution should be used for the OS curves for pembrolizumab plus axitinib and sunitinib. We note that the OS survival data is immature and therefore the long-term survival of patients treated with pembrolizumab plus axitinib is uncertain. The ERG considers that the same distribution should be used for both treatment arms and that the Weibull provides the best fit to five year survival data for sunitinib.
- **Time on treatment curves (ToT)**. The ERG considers that the same distribution should be used for all treatment arms and that the Weibull provides the best visual fit to the ToT data.
- **Age-adjusted utility.** The company found that there was no relationship between age and utility with the KEYNOTE-426 trial population (Clarification question B11). There was therefore no need to include age-adjusted utility.
- **Subsequent treatment costs.** Based on clinical advice to the ERG, we have modified the proportion of patients receiving subsequent treatments. For the pembrolizumab plus axitinib arm, more patients (20%) would receive cabozantinib.
- **Administration costs.** The cost of the administration of oral treatments was assumed to be £0, as in previous technology appraisals.
- **Terminal care costs.** The cost of terminal care was assumed to be £8,073, rather than £6,789.76, using the costs from the cabozantinib STA and updating to 2017/8.

The ERG preferred base case analysis gave an estimated ICER of £120,455 per QALY for based on list prices (Table 2).

Table 2 ERG base case cost-effectiveness for pembrolizumab + axitinib versus comparators in the overall population (pairwise comparisons)

Treatment	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental cost per QALY gained
Pembrolizumab + axitinib	[redacted]	[redacted]	-	-	-
Sunitinib	[redacted]	[redacted]	£140,895	1.170	£120,455
Tivozanib	[redacted]	[redacted]	£135,168	1.170	£115,558
Pazopanib	[redacted]	[redacted]	£137,335	1.170	£117,411

Subgroup analysis: intermediate / poor risk group

The ERG used the same set of preferred assumptions to estimate the ICERs for the intermediate / poor risk subgroup. The ERG preferred analysis gave an estimated ICER of £48,424 per QALY for pembrolizumab plus axitinib compared with cabozantinib (Table 3).

Table 3 ERG base case cost-effectiveness for pembrolizumab + axitinib versus comparators in the intermediate/poor risk subgroup (Pairwise comparisons)

Treatment	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental cost per QALY gained
Pembrolizumab + axitinib	[redacted]	[redacted]	-	-	-
Sunitinib	[redacted]	[redacted]	£141,941	1.010	£140,481
Tivozanib	[redacted]	[redacted]	£137,480	1.010	£136,065
Pazopanib	[redacted]	[redacted]	£139,200	1.010	£137,768
Cabozantinib	[redacted]	[redacted]	£44,012	0.909	£48,424

The ERG completed scenario analyses varying key assumptions in the model. For the overall patient population, the results vary between £72,591 - £162,424 per QALY gained for pembrolizumab plus axitinib compared to sunitinib. Those scenarios which have the largest effect on model results are changes to the distributions used for OS, using the log-logistic curve for ToT, including a waning effect and changes to the utility values.

For the intermediate/poor risk subgroup, the results vary between £27,892 - £149,347 per QALY gained for pembrolizumab plus axitinib compared to cabozantinib. Those scenarios which have the largest effect on model results are changes to the distributions used for PFS, using the log-logistic curve for ToT, including a waning effect, using time varying hazards (FP) and changes to the utility values.

1 Introduction to the ERG Report

This report is a critique of the company's submission (CS) to NICE from Merck Sharpe & Dohme on the clinical effectiveness and cost effectiveness of pembrolizumab in combination with axitinib for untreated advanced renal cell carcinoma. 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 the ERG via NICE on 8th August 2019. A response from the company via NICE was received by the ERG on September 2nd 2019 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

Section B.13 of the CS provides an overview of the key aspects of the aetiology and subtypes of RCC, its epidemiology and the clinical pathway of treatments including the proposed position of pembrolizumab plus axitinib. The ERG considers that the CS generally provides an accurate overview of RCC and its management. We summarise the key facts of relevance together with supplemental information where deemed appropriate below. The CS does not describe the impact of RCC on health-related quality of life (HRQoL) and we have provided a brief summary to highlight the potential impact RCC and its treatment has on an individual.

As stated in the CS, around 80% of all kidney cancer cases are RCC.[1] RCC typically originates in the lining of the tubules of the kidney; the tubules are responsible for filtering the blood and making urine. There are a number of subtypes of RCC according to the type of cells affected. The most common RCC subtypes are clear cell (75%), papillary or chromophilic (10-15%) and chromophobe (5%).[2]

RCC occurs more commonly in males than females and typically affects adults over 60 years.[3] The aetiology of RCC is unknown but risk factors include obesity, smoking and hypertension.[4] According to Cancer Research UK, statistics there are approximately 12,600 incidence cases of kidney cancer (no data specifically for RCC) each year (based on data from 2014-2016).[5] The CS states that incidence rates have increased rapidly (by 85%) since the early 1990s.[5]

RCC can be asymptomatic in the early stages and as such diagnosis can be made later in the disease process. In kidney cancer some 40% are diagnosed at a late stage.[5] The CS also states that in RCC around 44% presented at stage III or IV and that 25% to 31% had metastases;[5] the ERG note that these data are for kidney cancer generally rather than RCC. Initial symptoms that may be experienced by a person with RCC are haematuria (blood in the urine) and / or persistent lower back pain or pain between the ribs and hipbone.[3, 33]

RCC is typically staged from stage I to IV according to how far the cancer may have spread; stage III indicates that the cancer has advanced locally (within regional lymph nodes) and stage IV indicates that metastases beyond the regional lymph nodes are present (see Section 2.4 discussing this in the context of the NICE scope and CS decision problem).

Survival in RCC is linked to the stage of the cancer at diagnosis. In all kidney cancer cases the 1-year survival rate is 95% for those diagnosed at stage I compared with 37% for those diagnosed at stage IV.[5] Overall around 50% of people with kidney cancer live for at least 10 years.[5]

2.2 Symptoms and health-related quality of life

The CS does not describe the effect RCC can have in terms of symptoms or HRQoL. In a 2007 literature review of 12 studies, used as context for the development of a RCC symptom measure, the most commonly reported symptoms included fatigue, weakness, pain, anorexia, nausea and dyspnoea.[6] In a small (n=31) cross-sectional study that followed the literature review the five most reported symptoms in advanced RCC were fatigue, weakness, worry, shortness of breath and irritability.[6] In advanced RCC HRQoL is impaired by disease burden. The poor prognosis together with the symptoms associated with the disease can affect all domains of HRQoL including physical function, psychological factors such as depression and irritability, emotional status, sleep and social functioning[6, #38, #39] HRQoL improvements in advanced (metastatic) RCC have, however, been associated with tumour response, delayed progression and lower rates of adverse events from targeted treatments compared with previous treatments.[7, 39, 40]

2.3 Critique of company's overview of current service provision

The CS provides a limited overview of how advanced RCC is managed in UK clinical practice, summarising the NICE pathway for first-line treatment options (see CS Figure 2) and the European Association of Urologists (EUA) guideline for metastatic clear-cell RCC.[8] The EUA guidelines recommends first line pembrolizumab and axitinib as standard of care for people with International Metastatic RCC Database Consortium (IMDC) favourable risk disease (discussed in decision problem section below) except in those who cannot receive or are intolerant to immune checkpoint inhibitors (the class of drug that pembrolizumab belongs to). Pembrolizumab and axitinib is also recommended as a treatment option for people with IMDC intermediate or poor risk.[8]

2.4 Critique of company's definition of decision problem

The company's decision problem is as follows:

- Population: Adults with untreated advanced RCC.
- Intervention: Pembrolizumab in combination with axitinib
- Comparators: Tivozanib; pazopanib; sunitinib; cabozantinib (for disease that is intermediate- or poor-risk as defined in the IMDC criteria)
- Outcomes: overall survival (OS); progression-free survival (PFS); objective response rate (ORR); adverse events (AE); HRQoL.

The company's decision problem is largely consistent with the NICE scope. The population in the NICE scope is 'adults with untreated locally advanced or metastatic RCC'. The decision problem is 'untreated advanced RCC', and the ERG's clinical experts confirm that this can be taken to mean the same thing as on a practical level they both require systematic treatment. The CS decision problem also introduces a subgroup that was not noted in the NICE scope (there were no subgroups in the NICE scope). This was people with intermediate / poor risk category as defined by IMDC. The rationale for this addition is not made explicitly clear in the CS. The company's network meta-analysis states that the effect modification by RCC risk group is a justification for subgroup analyses (we discuss this further in section 3.1.7 of this report). The comparator treatment cabozantinib in the NICE scope is applicable only for this subgroup.

The CS summarises the IMDC risk evaluation in CS Table 4 as this was part of the stratification at randomisation for the participants in the pivotal phase III RCT KEYNOTE-426 (discussed in Section 3.1.3 of this report). Patients are assessed on six risk factors:

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

Participants are then placed in to one of three risk categories by totalling the number of risk factors: Favourable (0 factors); Intermediate (1 or 2 factors); Poor (3 or more factors).[9,10] Expert clinical advice to the ERG is that there is biological hypothesis for differential effect by prognostic risk and that the greater the risk the more likely patients will respond to an immune-oncology combination, than to tyrosine kinase inhibitor (TKI) monotherapy. Expert clinical advice also suggested that favourable risk patients have pre angiogenic characteristics with less tumour mutations, whereas intermediate/poor patients are less driven by angiogenesis, have more mutations, and higher PD-L1 expression. Thus, immunotherapy may be more effective in these patients than in favourable risk patients.

The company provides a summary description of pembrolizumab (but not axitinib) and its mechanism of action in CS Table 2. The ERG confirms this is consistent with the draft summary of product characteristics (SmPC, CS Appendix C). In addition to the differences between the NICE scope and the CS decision problem, the ERG notes that the evidence presented from the KEYNOTE-426 trial was not completely aligned with the decision problem. The trial population had locally advanced or metastatic RCC with clear cell component +/- sarcomatoid features (the latter refers to a highly aggressive form of RCC). The population may therefore not be generalisable to the wider RCC population (i.e. the estimated 25% without clear cell RCC). The ERG notes that the pivotal trials of the comparator treatments also comprised mostly or exclusively clear cell RCC patients. Previous NICE appraisals of these drugs did not restrict the patient population to those with clear cell RCC. Therefore, the current trial evidence is not out of line with that included in previous NICE appraisals. Additionally, there were some participants who had recurrent disease which may have been treated at the advanced stage.

3 CLINICAL EFFECTIVENESS

3.1 Critique of company's approach to systematic review

3.1.1 Description of company's search strategy

The literature search for clinical effectiveness studies is detailed in CS Appendix D. The search was designed to inform evidence submissions in a number of countries and thus had a broader scope than that of the current submission to NICE. Consequently, the strategy contains terms for treatments that are not used in the UK. The systematic review inclusion criteria were restricted to just those treatments included in the NICE scope. The search informs the company's network meta-analysis (NMA).

An appropriate range of databases was searched: Medline (including In-Process and other non-indexed citations); Embase and the Cochrane Central Register of Controlled Trials.
The search terms contain appropriate subject headings together with relevant free-text terms. A published search filter was used to identify RCTs. The sets are correctly combined and the number of hits (records retrieved) per line is documented for transparency. A combination of MeSH and free text terms were used. The description of the search process is transparent.

Supplementary searching was undertaken to identify ongoing trials on the National Institute of Health's (NIH) clinical trial registry (www.clinicaltrials.gov), the European Union (EU) Clinical Trials Register (www.clinicaltrialsregister.eu). Manual (by hand) searches of relevant conferences for the last two years was performed, including the American Society of Clinical Oncology (ASCO) and the European Society for Medical Oncology (EMSO). It is not stated whether reference lists of relevant studies were searched to identify further additional relevant studies.

The original database search was conducted in November 2018, and updated in February 2019. The ERG agrees that the company would be aware of all relevant RCTs for pembrolizumab plus axitinib (as stated on CS page 17). However, the ERG has run targeted searches for more recent evidence for the comparator trials included in the NMA. We used a citation pearl growing approach in Google Scholar on 16/08/19. We identified the pivotal trials of each of the comparators (cabozantinib[11], pazopanib[12] and tivozanib[13]) and used the characteristics of these articles to search for other relevant and authoritative materials. We cross-checked all studies

citing the three key publications for the trials in this indication against the studies identified by the CS. We also checked the clinicaltrials.gov database records for additional publications and ran a simple search on PubMed limiting studies to those published since February 2019. This search identified two unique references:

- An article in press which reports post hoc subgroup analyses of the COMPARZ trial of pazopanib versus sunitinib.[14] This article reports characteristics of pazopanib responders, patient subgroups who achieved better outcomes, and the effect of dose modification on efficacy and safety.
- An article in press which reports subgroup analyses of the CABOSUN trial of cabozantinib versus sunitinib (published ahead of print on August 9, 2019[15]

These articles do not appear to contain data relevant to the company's NMA and therefore we do not consider them any further in this report.

ERG conclusion:

The search strategies are comprehensive, well documented and are fit for purpose. It is unlikely that any potentially relevant studies of pembrolizumab plus axitinib and comparator treatments that were not identified and included.

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

A broad systematic literature review was conducted to meet the needs of multiple countries. The eligibility criteria for the broad review are not reported; a subset of the broader criteria is presented in CS Appendix D1.1 Table 1, which the company states reflect the target population for the UK NMA, and is discussed below. The aim was to identify studies to inform direct and indirect comparisons between the intervention and comparators of interest.

Population: the inclusion criteria specify adults diagnosed with histologically confirmed RCC with clear cell component with or without sarcomatoid features with the following staging:

- locally advanced (T3a–T4 per American Joint Committee on Cancer [AJCC])
- metastatic (Stage IV per AJCC), or chemo-naïve or chemo-experienced relapsed/recurrent disease of earlier AJCC stage.

These criteria reflect the eligibility criteria for the KEYNOTE-426 trial, but are narrower than the population specified by the NICE scope (RCC not limited to clear cell type).

Interventions and comparators: first line therapy with any of the following as monotherapy or combination therapy compared with each other or with placebo were eligible:

- Pembrolizumab plus axitinib
- Sunitinib
- Pazopanib
- Cabozantinib
- Tivozanib

These interventions are in line with the NICE scope (other than placebo, which is an appropriate comparator for inclusion in an NMA). In order to connect tivozanib to the network, the company included two interventions not relevant to UK practice. This meant inclusion of two trials that do not meet the eligibility criteria reported in CS Appendix D1.1 Table 1 (CS 2.9.3.1); see section 3.1.7 below.

Outcomes: eligible outcomes included those specified in the NICE scope (OS, PFS, response rates, adverse effects, HRQoL).

Study design: parallel group and cross-over RCTs, post-hoc subgroup analyses and open-label extension studies, and pooled analyses of RCTs (phase II and phase III) were eligible. Non-RCTs were excluded.

Other: Only studies published in English were eligible.

A flow diagram of study selection is presented in CS Appendix D1.1.2 Figure 1. The reasons for exclusion in the earlier stages reflect the criteria for the broader systematic review, with limits to first line treatment of clear cell RCC and interventions of interest to the UK scope occurring in the final stages. In response to clarification question C1, the company provided an updated flow diagram including numbers and reasons for exclusion of citations during the final stages (company clarification document Figure C1), although the ERG notes that this contains an error in the number of unique trials included/excluded. A total of 125 citations were excluded during the final stages: 67 were not relevant to first-line clear cell, and 58 had an intervention not of interest to the UK scope. A list of 728 studies excluded after full-text review of the broader review and reasons for exclusion is presented in CS Appendix D1.1.3 Table 9. This contains 70

references excluded for the reason ‘Intervention not relevant to UK perspective’; it is not clear why this is greater than the 58 stated in the flow chart.

ERG conclusion

The company has not been explicit about any potential bias in the selection of studies. Study selection was undertaken by two independent reviewers. Two trials of interventions not included in the decision problem were included, and it was not clear how these were identified, necessitating an ERG clarification question. It was explained that these were included to facilitate a connected network analysis.

3.1.3 Identified studies

One RCT, funded by Merck Sharpe & Dohme, compared pembrolizumab plus axitinib with one of the scoped comparators, sunitinib (KEYNOTE-426).[16] Two additional RCTs of the relevant comparators were also included for the NMA (see section 3.1.7 of this report).

Summary details of KEYNOTE-426 are presented in CS section B.2.2, including trial design, eligibility criteria, setting and locations, interventions, outcomes and statistical analysis (such as sample size and power, description of intention-to-treat (ITT) analysis). Details of participant flow are presented in Appendix D1.3 Figure 38. The trial journal publication,[16] protocol and clinical study report (CSR) were provided by the company.

KEYNOTE-426 included patients with previously untreated locally advanced/metastatic clear cell RCC, specified as newly diagnosed Stage IV RCC per American Joint Committee on Cancer or those with recurrent disease (CS page 21). The company clarified that Stage IV includes patients with T4, any N and M0, and any T, any N and M1 [ERG note: where T is size of tumour, N is lymph node involvement and M is distant metastases]. Those with T4, any N and M0 are considered as locally advanced as there is no metastatic disease. For those with recurrent disease, if disease recurred only within the renal fossa or with unresected kidney, this is also considered as locally advanced (clarification response A1).

As noted in CS section 1.3.1, clear cell RCC accounts for 75% of RCCs. Expert clinical opinion to the ERG is that type of RCC is not prognostic but describes a distinct clinical and biological entity. Almost all RCC treatment trials are conducted with patients with clear cell RCC and

require at least part of the tumour to have this histology. Other subtypes (e.g. papillary, chromophone) are different types of cancer but due to their rarity there are few trials in these disease subtypes. They are sometimes grouped together as 'non-clear cell RCC'.

The ERG noted that the baseline characteristics suggest some of the cases of recurrent disease may have been previously treated for stage III and IV disease, as 305 of the pembrolizumab plus axitinib arm and 328 of the sunitinib arm had stage III or IV at initial diagnosis and yet 238 and 231 were reported as having recurrent disease (CS Table 7). The company clarified that among participants with recurrent disease, 11 received adjuvant therapy and none received neo-adjuvant therapy. In the pembrolizumab plus axitinib arm, 4 participants had stage III RCC at initial diagnosis and received adjuvant therapy. None with recurrent disease and initial diagnosis of stage IV RCC received prior therapy (clarification response A2).

A total of 861 participants were randomised to receive:

- Pembrolizumab 200mg every 3 weeks by IV infusion for up to 35 doses (about 24 months) and
- Axitinib 5 mg twice daily orally (n=432)

or

- Sunitinib 50mg daily orally, 4 weeks on, 2 weeks off (n=429)

Treatments were continued until progressive disease was confirmed by blinded independent central review (BICR) or further confirmation by the investigator (details of this were provided by the company in response to clarification question A5 and are considered appropriate by the ERG), unacceptable adverse events, intercurrent illness preventing further administration of treatment, death or withdrawal of consent. If a participant was progression-free after 35 doses of pembrolizumab, treatment with axitinib was continued as monotherapy until progressive disease was confirmed. If either pembrolizumab or axitinib were discontinued because of toxicity or intolerance, treatment with the other drug was continued as monotherapy until progressive disease was confirmed. For both arms, if a complete response was observed and confirmed, treatment could be discontinued at the discretion of the investigator after a minimum of eight cycles of treatment (about 24 weeks) in the pembrolizumab plus axitinib arm or four cycles of treatment (about 24 weeks) in the sunitinib arm had been received. Retreatment (termed 'second course phase' in the CS) with pembrolizumab plus axitinib was permitted for up to 17

additional infusions of pembrolizumab therapy for participants who progressed after stopping treatment. The criteria for this were:

- Initial treatment with pembrolizumab stopped after a confirmed complete response and received at least 8 doses of pembrolizumab.

or

- Completed 35 doses (approximately 2 years) of pembrolizumab treatment without progressive disease.

and

- Investigator-confirmed radiographic disease progression after stopping initial treatment with pembrolizumab.
- No anti-cancer treatment since the last dose of pembrolizumab.
- Karnofsky Performance Status (KPS) of ≥ 70%.
- Adequate organ function.
- No history or current evidence of any condition, therapy, or laboratory abnormality that might interfere with participation for the full duration of the trial or is not in the best interest of the subject to participate.

The trial was undertaken in 16 counties. CS p.23 states 475 participants were enrolled in European sites (which equates to 55% of the 861 randomised) and this is confirmed on CS p.43 'Consideration of UK clinical practice', which states 55% of patients were recruited in Europe. However, CS Table 7 presents a lower proportion of 36.8% (317/861) of European participants. In response to clarification A6, the company states that the different proportions relate to the proportion enrolled (n=475, 55%) and the proportion after randomisation (n=317, 36.8%) and that some of the patients enrolled were not randomised. The ERG notes that based on these figures, the total sample size of the number enrolled is 861, the same as the total randomised (CS Appendix D1.3 Figure 38). The discrepancy is therefore not explained and the ERG considers that the higher proportion is misleading. CS p.23 states that 48 participants were enrolled in the UK, but the number randomised was not reported.

Baseline characteristics are presented in CS Table 7. These were balanced between groups and are summarised in Table 4. The median age of the trial population was 62 years (range 26 to 90 years) and 73% were men. About 80% had a KPS score of 90/100, and common metastatic sites were lung (72%), lymph node (46%) and bone (24%).

Table 4 Summary of baseline characteristics – KEYNOTE-426

	Pembrolizumab + axitinib (n=432)	**Sunitinib (n=429)**
Mean age (SD)	61.2 (10.0)	60.8 (10.2)
Median (range)	62.0 (30 to 89)	61.0 (26 to 90)
Male, %	71.3	74.6
White, %	79.4	9.5
Karnofsky Performance Scale, %		
90/100	80.3	79.5
70/80	19.4	20.5
Missing	0.2	0.0
IMDC Risk Category, %		
Favourable	31.9	30.5
Intermediate	55.1	57.3
Poor	13.0	12.1
PD-L1 Status, %		
CPS ≥ 1	56.3	59.2
CPS < 1	38.7	36.8
Not Available	0.9	0.5
Missing	4.2	3.5
Sites of Metastatic Disease, %		
Lung	72.2	72.0
Lymph Node	46.1	45.9
Bone	23.8	24.0
Adrenal Gland	15.5	17.7
Liver	15.3	16.6
Sarcomatoid Feature, %	11.8	12.6
Unknown, %	33.8	31.5
Missing, %	0.2	0.2
RCC Tumour Fuhrman Grade, %		
Grade 1, %	5.3	5.6
Grade 2, %	31.9	29.6
Grade 3, %	27.8	32.2
Grade 4, %	24.1	21.7
Missing, %	10.9	11.0
Disease Status at Baseline, %		
Recurrent	55.1	53.8
Newly Diagnosed	44.9	46.2
RCC Stage at Initial Diagnosis		
I	15.7	14.5
II	12.7	8.9
III	22.2	23.5
IV	48.4	52.9
Missing	0.9	0.2

IMDC, International metastatic RCC Database Consortium; PD-L1, program death-ligand 1; KPS, Karnofsky performance status; CPS, combined positive score; RCC, renal cell carcinoma.

Although typical of a phase III trial population, expert clinical advice to the ERG is that these patients are younger and fitter than the general population with untreated locally advanced/metastatic RCC.

Overall survival follow-up of the trial is ongoing, with an estimated completion date of January 2020.

ERG conclusion

The evidence for the clinical effectiveness of pembrolizumab plus axitinib versus one of the NICE scoped comparators, sunitinib, comes from one phase III RCT. The participants in the trial had previously untreated locally advanced/metastatic clear cell RCC. Other (non-clear cell) types of RCC, accounting for 25% of patients with RCC, were not included, but this is line with other trials of RCC. The majority (63%) of participants were from outside of Europe, and the number of participants randomised in the UK unclear.

3.1.4 Description and critique of the approach to validity assessment

The company provided a quality assessment of KEYNOTE-426 using the Cochrane Risk of Bias criteria (version 1). A comparison of the company and ERG assessments is presented in Table 5. The ERG has also completed additional questions from the NICE recommended quality criteria. The ERG generally agrees with the company's judgements, and where there are disagreements the ERG's judgements are in a more positive direction. The ERG considers that the risk of selection bias due to inadequate concealment of allocations prior to assignment was low due to the use of a central interactive voice response system /integrated web response system. Also, while the open-label design of the trial meant that there was a high risk of performance bias due to knowledge of the allocated interventions by participants and personnel, the risk of detection bias was low due to the use of blinded central review of progression and response. The groups were similar at baseline in terms of prognostic factors, and appropriate intention to teat analysis and methods to account for missing data were used.

Table 5 Company and ERG assessment of trial quality for KEYNOTE-426

	CS response	ERG response
Cochrane risk of bias domain		
Sequence generation	Low risk	Low risk
Allocation concealment	Unclear risk	Low risk (central randomisation using an interactive voice response system /integrated web response system
Blinding of participants, personnel and outcome assessors	High risk	High risk for participants and personnel Low risk for outcome assessors
Incomplete outcome data	Low risk	Low risk
Selective outcome reporting	Low risk	Low risk
Other sources of bias	Low risk	Low risk
Additional NICE quality assessment criteria		
Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Not reported	Yes – low risk
Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Not reported	Yes – low risk

[a] Low = low risk of bias, high = high risk of bias, unclear = uncertain risk of bias.

ERG conclusion

There is a risk of performance bias in the trial, but the ERG considers the trial to have a low risk of bias for the other domains.

3.1.5 Description and critique of company's outcome selection

All outcomes specified by the NICE scope [overall survival, progression-free survival (PFS), response rates, adverse effects of treatment and health-related quality of life (HRQoL)] were measured in KEYNOTE-426 and are presented in the CS. Other relevant outcomes assessed in KEYNOTE-426 included time to deterioration, duration of response (DOR) and disease control rate (DCR); these are presented in CS Appendix L.

Overall survival and PFS were dual primary endpoints in the trial. Overall survival was defined as the time from randomisation to death due to any cause, with censoring at the date of the last-follow-up for participants without documented death at the time of final analysis (see section

3.1.6). PFS was defined as the time from randomisation to the first documented disease progression per Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 criteria assessed by blinded independent central review (BICR) or death due to any cause. PFS based on investigator assessment (RECIST 1.1) is presented in CS Appendix L Table 3 (the ERG notes the heading of this table states BICR, however we have checked against the CSR that these are indeed investigator assessment). The investigator assessments produced similar results that were also statistically significant, although the HR was slightly larger (less favourable towards pembrolizumab plus axitinib), see section 3.3 of this report. The ERG considers the BICR assessment to be have a lower risk of bias than the investigator assessment.

The following outcomes were secondary endpoints. Objective response rate (ORR) was defined as the proportion of participants with a complete response (CR) or partial response (PR). For those with CR or PR, DOR was defined as the time from first documented evidence of CR or PR until disease progression or death due to any cause. DCR was defined as the proportion who achieved CR, PR or stable disease (SD) for ≥ 6 months. Assessments were by BICR according RECIST 1.1 criteria. ORR, CR and PR based on investigator assessment (RECIST 1.1) are not presented in the CS but are available in the CSR; the ERG notes that these were similar to BICR results, see Results section 3.3.2. DOR and DCS based on investigator assessment are presented in CS Appendix L Tables 7 and 10, see Results section 3.3.2.

HRQoL was assessed by longitudinal score changes from baseline in the global health status/quality of life scale of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 items (EORTC QLQ-C30), a validated PRO measure. The ERG is aware that minimal clinically important differences have been identified for a population with advanced cancer (including renal cancer); these vary for each of the scales and symptoms.[17] The trial protocol also states that the proportions of people with deterioration/stable/improvement at 42 weeks (based on expected median PFS of 11 months in the control group) into the study will be described, however this is not presented in the CS (proportions at 30 weeks are available in the CSR). Supportive analyses of five functional scales (physical, role, emotional, cognitive and social), three symptom scales (fatigue, nausea/vomiting and pain) and six single items (dyspnoea, sleep disturbance, appetite loss, constipation, diarrhoea and financial impact) were also undertaken, however only selected scales are presented in CS Appendix L and the CSR (physical functioning, role functioning,

nausea/vomiting, diarrhoea) (the ERG requested results for all scales, clarification question A7, and these were subsequently provided by the company).

Utility was measured using the EQ-5D-3L as an exploratory endpoint; this is a standardised and validated generic instrument and is NICE's preferred measure of HRQoL in adults. EQ-5D-3L utility data at baseline and end of trial for each arm of the trial are not presented in the CS or related appendices; only the EQ-5D visual analogue score (VAS) is presented. This is a qualitative measure of health reflecting the patient's own judgement, on a scale from 'best imaginable health state' to 'worst imaginable health state'. On request from the ERG the company provided the EQ-5D-3L index data at baseline and week 30 for each treatment (clarification question A8).

The PROs were completed prior to all other study procedures and were assessed on Day 1 of each cycle in the pembrolizumab plus axitinib group, and on Days 1 and 29 of each cycle up to Cycle 4, then on Day 1 of each subsequent cycle following the two-week-off treatment period in the sunitinib group. 'Compliance and completion rates' (as one outcome) for the FKSI-DRS, EORTC-QLQ-30 and EQ-5D-3L at baseline and at week 30 are reported in CS Appendix L Table 14. Compliance rates appeared to be slightly lower in the pembrolizumab plus axitinib arm at both baseline and week 30. In both treatments arms, rates decreased between baseline (pembrolizumab plus axitinib: approx. 92%, sunitinib approx. 97%) and week 30 (pembrolizumab plus axitinib: approx. 86%, sunitinib approx. 89%).

Adverse events and serious adverse events were graded according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.0. Drug-related adverse events were determined by the investigator to be related to the drug.

Overall survival, PFS, utilities from EQ-5D-3L, and Grade ≥3 all-cause adverse events occurring occurred in at least 5% of patients are used in the economic model.

ERG conclusion

The ERG considers that the outcomes included in the CS are appropriate and relevant to the NICE scope.

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

3.1.6.1 Sample size calculation and hypotheses

The trial was designed to test the superiority of pembrolizumab plus axitinib vs sunitinib with respect to PFS and OS, the co-primary outcomes.

The sample size calculation was conducted for the two primary outcomes of the study, PFS and OS. The power calculation was based on the final number of randomised patients (n=861, an increase on the original power calculation based on 840 patients).

- Expected median PFS in the sunitinib arm was 13 months. Based on 487 required PFS events the trial had approximately 99% power to detect a HR of 0.60 for PFS (pembrolizumab plus axitinib vs sunitinib) at alpha=0.2% (1-sided).
- Expected median OS in the sunitinib arm was 33 months. Based on 404 required death events, the study had 80% power to detect a HR of 0.75 for OS at alpha=2.3% (1-sided).

The CS states that the assumptions for PFS and OS were based on emerging data from the sunitinib arm of the CheckMate 214 trial of nivolumab plus ipilimumab.[18]

3.1.6.2 Data analysis timepoints

Three analysis timepoints were planned, two interim and one final.

Interim analysis 1 (IA1) – the first interim analysis for PFS and OS, after completion of enrolment and a minimum 7-month follow-up. The minimum expected PFS events was n=305; the required OS events was n=195, or 48% of expected number. Data cutoff was the 24th August 2018. Median duration of follow-up at this time was 13.2 months (pembrolizumab plus axitinib) and 12.1 months (sunitinib). The study showed statistically significant improvement in both co-primary endpoints and key secondary endpoint. Efficacy testing was therefore stopped at IA1.

Interim analysis 2 (IA2) – the final analysis for PFS (n=487 events expected) and the second interim analysis for OS (n=299 events expected, 74%). However, this analysis was not conducted. Instead, a second (unplanned) data cut was taken in January 2019 for US Food and

Drug Administration (FDA) regulatory purposes (Safety Update Report). The statistical significance of all pre-specified tests was already achieved at IA1, so this analysis provides no further formal hypothesis testing results. Results are presented in CS Appendices F, N and O. These are not used in the assessment of cost-effectiveness in the CS.

Final analysis – the final planned analysis for OS (if not already declared successful), to take place when 404 events have occurred.

The CS does not mention any implications of the early stopping of the trial analysis at interim analysis 1 on the estimation of the size of the treatment effect. The ERG notes that there have been debates in the literature about the impact of early stopping of trials on the effect estimates.[19] Early stopping can sometimes result in over-estimation of treatment effect. A simulation study showed that in trials with a well-designed interim-monitoring plan, stopping the trial when 50% or greater of the information has been collected has a negligible impact on estimation.[20] A total of 395 PFS events had been recorded at this time, which is 81% of the total events required overall (n=395/487). Thus, this outcome is unlikely to have been over-estimated. However, for OS only 156 events had been recorded, which is 39% of the overall total required (n=156/404). Thus, the available OS results at the interim analysis are potentially over-estimated and should be interpreted with caution due to immaturity.

3.1.6.3 Analysis populations

Two statistical analysis populations are included in the trial:

- Intention to treat (ITT) population (n=861/861 randomised, 100%), defined as all randomised patients included in the trial. Patients were analysed in the treatment group to which they were randomised. This ERG regards this as an appropriate way of conducting ITT analysis. The ITT population was used for the analysis of efficacy outcomes.
- All Subjects as Treated (ASaT) population (n=854/861 randomised, 99%). This consisted of all randomised patients who received at least one dose of study treatment. Patients were analysed in the treatment group corresponding to the study treatment they received. This population was used for the analysis of safety outcomes.

The ERG considers the definitions of these two analysis populations appropriate.

3.1.6.4 Disease progression assessment

CS Table 10 summarises the primary censoring rule used, and variations of the rules explored in sensitivity analysis (results of these given in the CS Appendix L).

The primary censoring rule for death or disease progression was estimated by the date of the first assessment at which progressed disease was objectively documented (per RECIST criteria 1.1) by blinded independent central review (BICR). Patients who did not experience a PFS event were censored at the last disease assessment. Any patients who commence new anti-cancer therapy were censored at the last disease assessment prior to the initiation of new anti-cancer therapy.

Three further potential censoring scenarios are proposed, based on different possibilities on the number and timing of missed disease assessments are proposed (CS Table 10). The primary censoring rules and sensitivity analyses associated with these scenarios are stated. The range of scenarios explored and assumptions about when progression occurred appears to be comprehensive.

Sensitivity analyses were performed for the comparison of PFS based on investigator's assessment and PFS with progressive disease as determined per RECIST by immune-related BICR. Results of the sensitivity analyses are reported in CS Appendix L.

The CS reports that the proportional hazards assumption for PFS could be examined using both graphical and analytical methods. No information is reported regarding proportionality for OS. See section 3.1.7.5 of this report for further discussion of the proportional hazards in relation to the NMA and section 4.3.5 in relation to the survival curves used in the economic modelling.

3.1.6.5 Subgroup analyses

Pre-specified subgroup analyses were performed to determine whether the treatment effect was consistent across the following subgroups:

- IMDC risk category (favourable versus intermediate versus poor; favourable versus intermediate plus poor)
- Geographic region (North America versus Western Europe versus Rest of the World)
- PD-L1 status (combined positive score [CPS] <1 versus CPS ≥ 1)

- Age (< 65 versus ≥ 65)
- Sex (male versus female)
- Race (white versus non-white)

Results are presented in the CS Appendix E for PFS and OS and additionally for ORR in the clinical study report. The ERG notes that results for two additional subgroups are presented for these outcomes, Karnofsky performance scale score at baseline (90/100; 70/80) and number of metastatic organs (1; ≥2). These are not mentioned as being pre-specified subgroups in the CS, or trial protocol, so the ERG assumes that these were included post-hoc.

The ERG considers that these chosen subgroups are appropriate to this disease and we are not aware of any other key subgroups that have been omitted. The interpretation of the results of the subgroup analyses should be made with caution as the number of patients in some subgroups is relatively small. Further caution is required for subgroup analyses by OS, as data for this outcome is currently immature.

The ERG asked the company to specify whether any statistical interaction tests were conducted during the subgroup analyses (clarification question A4). The company responded that there were no pre-specified interaction tests performed for subgroup analyses in the trial because at the study design stage, an interaction effect between subgroups was not expected. Statistical interaction tests can confirm statistically significant differences in effect between subgroups of interest, if detected.

3.1.6.6 Statistical tests used

The non-parametric Kaplan Meier (KM) method was used to estimate the PFS and OS curves in each treatment group.

The statistical tests used for PFS and OS were the stratified Logrank test estimation. A stratified Cox proportional hazard model with Efron's method of tie handling was used to assess the magnitude of the treatment difference (HR) between the treatment groups. The stratification factors were those used in the randomisation process (i.e. IMDC risk status, and geographic region). The ORR was assessed by the stratified Miettinen and Nurminen method.

3.1.6.7 Multiplicity in statistical testing

Running multiple statistical tests increases the probability of finding statistically significant results by chance even if there is no underlying effect. A pre-specified multiplicity strategy was applied to the two primary outcomes, PFS and OS, and the secondary outcome ORR. The strategy was based on the approach of Maurer and Bretz (CS Figure 4). In summary, the Type I error (α) allocated to a hypothesis that was successfully tested is re-distributed for testing of the other two hypotheses.

Initially, α=2.3% (23/25 of the overall total α = 2.5% for testing the OS, PFS and ORR) is allocated to the OS hypothesis and α =0.2% (2/25 of the overall total α =2.5%) is allocated to the PFS hypothesis. A series of steps was then followed whereby if the OS null hypothesis was rejected, half of its α was reallocated to PFS testing, and if the null hypothesis for OS and ORR were both rejected all α's were reallocated to the PFS hypothesis test. Similar steps were followed for the testing of the OS hypothesis. The ORR hypothesis was initially allocated a Type I error α =0% and thus could be tested unless one or both PFS or OS null hypotheses were rejected. Full details of the multiplicity strategy can be found in CS section 2.4.1.

The ERG considers the procedures followed in the trial to prevent statistically significant effects being detected by chance to be appropriate.

3.1.6.8 Analysis of safety

The CS reports that the safety analyses used a tiered approach. The tiers differed with respect to the analyses that was performed. Tier 2 parameters were assessed via point estimates with 95% CIs provided for between-group comparisons; only point estimates by treatment group are provided for Tier 3 safety parameters. The 95% CIs for the between-treatment differences in percentages are provided using the Miettinen and Nurminen method. There were no Tier 1 events in this trial. The CS does not define which events would be classified according to which tier. However, this information can be found in the trial protocol.

ERG conclusion

The statistical analyses in the KEYNOTE-426 trial are appropriate for the evaluation of a cancer therapy, and appear to have been implemented correctly. The trial was adequately statistically powered for the primary efficacy outcomes; procedures were

used to address potential multiplicity in statistical hypothesis tests; an appropriate ITT analysis was conducted; appropriate censoring rules for assessing PFS were used (with sensitivity analyses to examine robustness of the censoring approach); and appropriate pre-specified subgroup analyses were conducted.

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

As only one trial of pembrolizumab plus axitinib in this indication was included in the submission (KEYNOTE-426), a meta-analysis of pembrolizumab trials was not possible. The CS provides a narrative review of the trial, with data presented in tables and text. As the only head-to-head comparison available was between pembrolizumab plus axitinib versus sunitinib it was necessary to conduct indirect comparisons to the other treatments in the decision problem. The company uses network meta-analyses (NMA) for this purpose.

3.1.7.1 Overview of network meta-analysis (NMA) approaches used

The CS reports two types of Bayesian approaches for indirect comparison of pembrolizumab plus axitinib with other treatments:

- Network meta-analysis (NMA) assuming constant hazards
- NMA assuming time-varying hazards based on fractional polynomials.

Both of these NMAs assess OS and PFS outcomes, but not the other outcomes relevant to the decision problem (response rate, HRQoL, adverse events).

The NMA assuming constant hazards appears to be the 'primary' indirect comparison method reported in the CS. The results of this analysis are reported in CS section 2.9.3. Details of the NMA using fractional polynomials are largely reported in appendices (CS appendix D and M).

It should be noted that these NMAs do not inform all estimates of cost effectiveness in the economic model. For the base case economic analysis the economic model uses patient-level data on OS, PFS and safety from the KEYNOTE-426 trial, with pazopanib and tivozanib assumed to be clinically equivalent to sunitinib. In section 4.3.5 of this report we give further detail on the clinical effectiveness estimates used in the economic model, and we note that the sunitinib PFS estimates from KEYNOTE-426 are in line with previous pivotal trials of sunitinib.

The NMA results do inform the economic model for the comparison of pembrolizumab plus axitinib and cabozantinib. As will be explained below, this analysis was restricted to patients at intermediate/poor RCC risk and is analysed separately as a sub-group analysis.

It should also be noted that in previous NICE appraisals of first line treatments for advanced RCC the appraisal committees have agreed, based on expert clinical opinion, that sunitinib, pazopanib and tivozanib are broadly similar to each other in efficacy and safety, and therefore have not considered indirect comparisons as a key factor in their decision making (see Appendix 9.1). However, the current appraisal includes cabozantinib as a comparator, and this has not been directly compared against pembrolizumab plus axitinib.

Notwithstanding potential judgements about the necessity of an NMA in the current appraisal, we have conducted a critique of the NMA, detailed in in the following sub-sections. A summary of the NMA results are presented in section 3.3.5 of this report.

3.1.7.2 Evidence networks

CS section B.2.9.1 provides an overview of the evidence networks constructed for the NMA. The company's systematic review of clinical effectiveness identified four RCTs evaluating five treatments relevant to the decision problem and inclusion in the NMA (CABOSUN,[11], COMPARZ,[12] KEYNOTE-426,[16] and TIVO-1[21]). A further two trials of treatments not included in the decision problem are included in the NMA to allow tivozanib to be connected to the network (see below for a discussion of these trials).[22, 23]

The networks are presented as a base case analysis, which included all patients irrespective of baseline RCC risk status, and subgroup analyses for patients at intermediate/poor RCC risk, and patients at favourable RCC risk.

Base case analysis

A visual representation of the base case network for all included RCTs and for all outcome measures is provided in CS Figure 10, reproduced below in Figure 1 (CS Figure 10), and tabulated in Table 6. This applies to the constant hazards NMA and the time-varying fractional polynomials NMA.

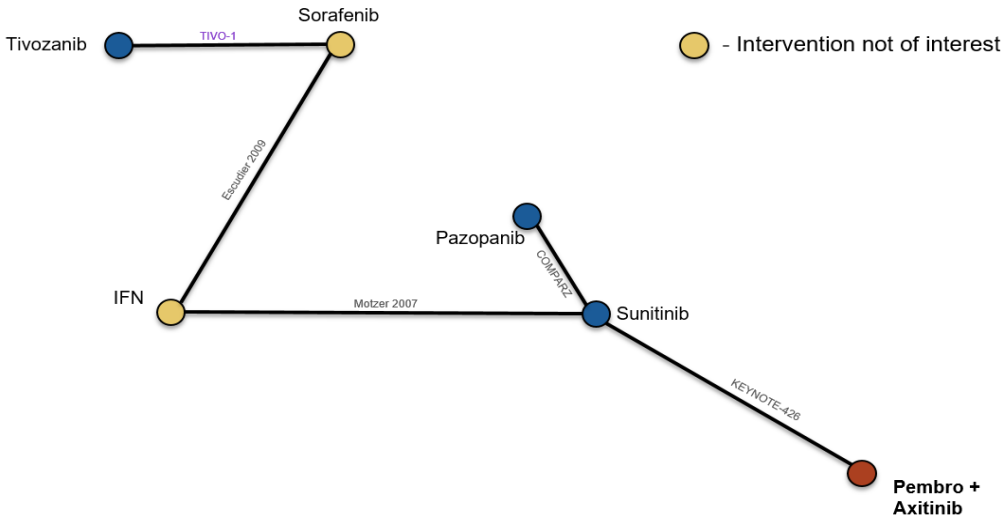

Figure 1 Network of RCTs in the base case NMA (all outcome measures)

Reproduced from CS Figure 10
NB. The CABOSUN trial (cabozantinib vs. sunitinib) is not included in this network diagram as this trial included IMDC intermediate/poor risk category patients only

Table 6 Summary of RCTs included in the NMA

Trial identifier	Intervention A	Intervention B
CABOSUN[11]	Cabozantinib	Sunitinib
COMPARZ[12]	Pazopanib	Sunitinib
KEYNOTE-426[16]	Pembrolizumab + axitinib	Sunitinib
TIVO-1[13]	Tivozanib	Sorafenib
Escudier et al[a]	Sorafenib[a]	Interferon alpha[a]
Motzer et al[a]	Sunitinib	Interferon alpha[a]

[a] Intervention not relevant to the decision problem. Included in network to connect relevant treatments together

The base case includes six RCTs evaluating four of the five treatments relevant to the decision problem (pembrolizumab and axitinib, sunitinib, pazopanib, tivozanib). The fifth treatment, cabozantinib, is only included in the intermediate/poor risk subgroup (see below). Two treatments not included in the decision problem (interferon alpha and sorafenib, from two RCTs[22, 23]) were only included to allow tivozanib to be connected to the network. The CS does not report how these two trials were identified, whether from the company's systematic review of clinical effectiveness, or another source. The ERG asked the company to clarify how these studies were identified and selected, and whether there were any other relevant studies which

could have been used (clarification question A9). The company responded that the two studies were identified from a broader systematic literature review they conducted for health technology assessments in multiple countries (NB, the systematic review of clinical effectiveness for the current appraisal was a subset of this broader systematic review, and included UK relevant treatments only). The company states that based on their latest literature search (conducted in February 2019), no other studies were identified that could be used to facilitate a connection of tivozanib to the evidence network. The ERG has checked the list of studies excluded from the systematic review of clinical effectiveness (CS appendix D1.1.3) and previous NICE appraisals of first line treatment for advanced RCC, and is not aware of any other relevant trials that could have been included in the NMA to connect tivozanib to the network.

Outcome-specific networks are depicted in CS Figure 11 (PFS) and Figure 12 (OS). The OS network contains fewer trials and relevant treatments than depicted in Figure 1 (CS Figure 10) due to the lack of available HR and Kaplan-Meier data needed for the constant HR and the time-varying hazard analyses, respectively. Notably, tivozanib is not included in the OS network since the connecting study Escudier 2009 did not report OS. Table 7 provides an overview of which treatments were included in the NMA base case and subgroup analyses (constant hazard and time-varying hazard fractional polynomial), by outcome measure.

Table 7 Treatments included in the NMA base case and subgroup analyses (constant hazard and time-varying hazard fractional polynomial), by outcome measure

Treatment	Base case		Intermediate/poor risk subgroup		Favourable risk subgroup[a]	
	PFS	OS	PFS	OS	PFS	OS
Pembrolizumab + axitinib	✓	✓	✓	✓	✓	✓
Sunitinib	✓	✓	✓	✓	✓	✓
Pazopanib	✓	✓	N/A	N/A	✓	✓
Tivozanib	✓	×	N/A	N/A	×	×
Cabozantinib	N/A	N/A	✓	✓	N/A	N/A

✓ = analysis conducted, × = analysis not possible
N/A = analysis not applicable
[a] constant hazard NMA only

Note that the TIVO-1 trial enrolled patients to receive first-line or second-line treatment. Therefore, subgroup data, rather than the data from all randomised patients, from this trial were used in the NMA. NMA results for TIVO-1 should be interpreted with caution as randomisation has been broken for this trial.

The CS reports that patient crossover occurred in one trial, TIVO-1 (in which 62% of patients in the sorafenib arm of the trial switched to tivozanib on disease progression) but noted that "outcomes included in the analyses described herein do not include patients who crossed over" (CS Appendix D.1.2). However, this is at odds with the company's response to clarification question A19 which states "The data used in the NMA does not account for cross-over, therefore, OS results where TIVO-1 is included is confounded by cross-over. Crossover was allowed only for patients who progressed on sorafenib to receive tivozanib, which may confound OS" (clarification question A19 response). The previous NICE appraisal committee for TA512 (tivozanib) were critical of the NMA for not adjusting for crossover in the included trials. The committee concluded that not adjusting for crossover meant that the results of the NMA were likely to be confounded with the direction of bias unknown. However, we do not believe this to be an issue in this current appraisal as OS from tivozanib is not included in the NMA. A crossover-adjusted HR for OS (using inverse probability of censoring weights method) was available in the tivozanib technology appraisal guidance (tivozanib vs sorafenib HR 1.02, 95% CI 0.67, 1.55) but could not have been used in the analysis as the "connecting" trials did not report OS. Similarly, the Escudier 2009 study also permitted crossover but only pre-crossover results are included in the NMA. Despite the apparent contradictory statements in the CS and clarification response, since no further connecting studies could be found the ERG considers no further analyses or adjustment necessary.

Subgroups

The NICE scope for this appraisal did not specify any subgroups of relevance. However, the company conducted separate NMAs for RCC risk subgroups: intermediate/poor and favourable. As a justification for these analyses the CS states that RCC risk score is an effect modifier in the treatment of RCC. The ERG notes that the subgroup analysis of the KEYNOTE-426 trial did not report statistically significant subgroup interactions by RCC risk group. As mentioned earlier in this report, the company stated in response to an ERG clarification question that there were no pre-specified interaction tests performed for subgroup analyses in the trial because at the study design stage, an interaction effect between subgroups was not expected.

Expert clinical advice to the ERG is that RCC risk status is an important prognostic factor. The experts also commented that there is empirical evidence that it is an effect modifier, as demonstrated in a recent phase III RCT of ipilimumab plus nivolumab, the Checkmate 214 RCT,[18] which was designed to show treatment effect differences according to risk groups. The

ERG asked the company to provide evidence to back up the assertion of effect modification for these risk subgroups (clarification question A20). In response, the company stated that risk status is a recognised prognostic factor in RCC and thus was therefore considered a potential treatment effect modifier. They cite the COMPARZ and CABOSUN trials as empirical evidence of this.

The ERG agrees with the company's decision to conduct a separate NMA for the intermediate/poor RCC risk group, as inclusion in the CABOSUN trial was restricted to patients in these risk groups, and cabozantinib is recommended by NICE only for patients at intermediate/poor risk (as defined by the IMDC criteria) (NICE TA542[24]).

As can be seen from Table 7 (and Figures 7 to 10 in CS Appendix D), the IMDC intermediate/poor risk network includes three treatments from two RCTs: pembrolizumab plus axitinib versus sunitinib (KEYNOTE-426), and cabozantinib versus sunitinib (CABOSUN). Pazopanib and tivozanib are missing from the network since no relevant subgroup data were reported. The intermediate/poor risk patients comprise the whole randomised population in the CABOSUN trial, but they are a subgroup of the KEYNOTE-426 trial (n=592/861;69%).

The favourable risk RCC NMA subgroup comprised three treatments (from two RCTs – KEYNOTE-426 and COMPARZ) Table 7 (and Figures 11 to 14 in CS Appendix D): pembrolizumab and axitinib; sunitinib and pazopanib. It was not possible for the company to conduct a time-varying fractional polynomial NMA in this subgroup as only one trial reported the necessary Kaplan Meier data necessary (KEYNOTE-426). This is the only network where both a constant hazard and a fractional polynomial model could not be conducted. The CS does not report separate cost-effectiveness estimates for patients with favourable RCC risk.

The ERG notes that the CABOSUN trial showed differences in PFS between intermediate and poor risk groups: in both groups the HR favoured cabozantinib over sunitinib, however in the poor risk group the confidence interval was wide and crossed one (likely due to small subgroup sample size). The ERG asked the company to run separate NMA scenario analyses for (i) intermediate risk patients, and (ii) poor risk patients (clarification question A21). We summarise these results in section 3.3.5 of this report, and our overall conclusion is that it cannot necessarily be concluded that there are differences in effect (for PFS) between poor and intermediate risk subgroups.

Overall, caution is required in the interpretation of the subgroup NMA analyses. They are based on a subset of randomised patients from the KEYNOTE-426 trial, and this can increase uncertainty about the precision of treatment effects.

3.1.7.3 Clinical heterogeneity assessment

CS Appendix D1.2 details the characteristics of the included trials (Table 10, Table 12, Tables 14 to 17; Figures 15 to 37). Details of the two trials of treatments not included in the decision problem[22,23] (interferon alpha and sorafenib) included to connect tivozanib to the network are provided in the company's response to ERG clarification question A9.

All trials were phase III RCTs, except CABOSUN and the trial by Escudier et al which were both phase II trials. They ranged in sample size from 157 patients (CABOSUN) to 1110 (COMPARZ). The trials were similar in terms of: inclusion criteria; sunitinib dosing schedule (where applicable); patient demographic characteristics (age, gender, race/ethnicity – where reported) and prior radiotherapy treatment.

The trials were generally similar in terms of the proportion of patients with lung, bone, liver and lymph node metastases. However, the CABOSUN trial and the trial by Escudier et al had a slightly higher proportion of patients with bone metastases (around 35% compared to 15% to 24% amongst the other trials). The ERG report for the cabozantinib NICE appraisal (NICE TA542) notes that the cabozantinib CS states that patients with bone metastases have a poor prognosis and experience poorer health outcomes with currently available treatments compared with patients without bone metastases. The current CS does not mention this. Expert clinical advice to the ERG states that bone metastases has a worse prognosis and can pose management problems. One expert commented that it may not be essential to consider evidence in patients with bone metastases separately from evidence in patients without.

In terms of baseline cancer performance score, four trials reported the distribution of Eastern Cooperative Oncology Group (ECOG) scores of the patients (CABOSUN, Escudier et al, Motzer et al; TIVO-1), and two reported Karnofsky scores (COMPARZ and KEYNOTE-426). Expert clinical advice to the ERG is that the Karnofsky scale (from which the ECOG is derived) is less commonly used but its scores can be mapped to ECOG scores to assess comparability. The majority of patients across the trials were classed as either ECOG 0 or 1 (meaning the patient is

able to function day to day without serious restriction), or Karnofsy score 90 to 100 (able to carry on normal activity and work; no special care needed). Only the CABOSUN trial included ECOG 2 patients (defined as ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours around (13% of patients). This is likely because this trial included only patients at intermediate or poor RCC risk, and ECOG performance status is one of the constituent variables in the IMDC risk status assessment. With this exception, overall the ERG concludes that the trials can be considered similar in terms of patient cancer performance status.

RCC risk status was measured by IMDC criteria in two trials (CABOSUN; KEYNOTE-426) and by MSKCC in four trials (COMPARZ and TIVO-1). The COMPARZ trial also assessed risk according to prognostic criteria by Heng et al[9], which was the basis of the later IMDC risk criteria [10]. As already discussed, expert clinical advice to the ERG is that the IMDC and MSKCC risk criteria can be considered similar. Thus, differences between the trials in the risk status patients were classified as would be unlikely. With the exception of CABOSUN trial and the trial by Escudier et al, the trials were similar in the distribution of patients across risk categories: 27-35% favourable risk; 55-64% intermediate risk, 5-13% poor risk. As already noted above, CABOSUN only included patients at intermediate or poor risk, with the proportion of randomised patients in these categories at 81% and 19% respectively. Escudier et al included a greater proportion of patients at favourable risk (52%) and intermediate risk (47%), with only 1% (a single patient) at poor risk.

There was some variation between the trials in the proportion of patients receiving prior nephrectomy. In one trial prior nephrectomy was an inclusion criterion (TIVO-1). With the exception of CABOSUN, in the other trials the proportion ranged from 94% (Escudier et al) to 83% (COMPARZ and KEYNOTE-426). The proportion was lowest in CABOSUN (75%). 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, hence why nephrectomy was lower amongst intermediate/poor risk patients in CABOSUN. Expert clinical advice also notes that prior nephrectomy may be associated with a better treatment outcome, thus raising the potential risk of bias in the NMA results. However, expert advice suggested that evidence for this is contradictory and it is an issue undergoing debate at scientific conferences.

The ERG is not aware of any key prognostic factors or effect-modifying characteristics that differ between the included trials. Expert clinical advice to the ERG agrees.

ERG conclusion

The trials included in the NMA can be considered similar to each other in terms of patient demographic and prognostic characteristics and in sunitinib treatment regimens. An exception to this is the CABOSUN trial which differed from the other trials on a number of characteristics, as outlined above (e.g. phase II trial, smaller sample size; only included patients at intermediate or poor RCC risk; higher proportion of patients with bone metastases; lower proportion of patients receiving prior nephrectomy; inclusion of patients with ECOG 2 performance status). These differences are likely to be an artefact of the intermediate/poor risk status of the patients in this trial. The impact of these differences on the results of the NMA are mitigated by exclusion of CABOSUN in the base case NMA. Instead, it is included in a subgroup analysis of patients with intermediate/poor risk.

3.1.7.4 Critical appraisal of trials included in the NMA

CS Appendix D.1.2.5 provides a risk of bias assessment of the four trials included in the NMA, using Cochrane risk of bias criteria (version 1). The CS considers that, overall, the trials were considered to have low risk of bias, aside from bias associated with open-label trials. The ERG conducted an independent risk of bias assessment of the trials (Appendix 9.2) and mostly agreed with the appraisal judgements of the company, with some exceptions. These exceptions tended to be where the ERG considered the risk of bias to be unclear, rather than low or high. The ERG concurs with the company's overall conclusions that the trials were at low risk of bias, except for bias arising from lack of blinding.

3.1.7.5 Proportional hazards assumptions

Indirect comparisons of time-to-event data are generally made using the assumption of proportional hazards. Where the proportional hazards assumption is not supported alternative approaches can be used, based on the assumption of time-varying hazards. The *a priori* rationale for using both constant hazards and time-varying hazards NMA assumptions in the CS is not explicitly stated. The CS provides cumulative hazard plots and log cumulative hazard plots for KEYNOTE-426 (CS section B.3.3). The log-cumulative hazard plots for OS in the trial

cross suggesting the proportional hazard assumption does not hold. Additional tests of proportional hazards, such as Schoenfeld residual, were not presented in the CS. For other trials in the NMA, only Kaplan Meier plots were available to the company to inform assessments of proportional hazards.

In CS Section B 2.9.4 the company discuss assumptions of proportional hazards based on a comparison of the results of the constant hazards and the time-varying hazards fractional polynomial NMAs. The company concluded proportional hazards did not hold for PFS (base case) and OS (intermediate/poor subgroup) (Table 8) However, in response to ERG clarification question A16 the company maintained that despite the violation of proportional hazards assumption, the fractional polynomial models are "more sensitive and detect chance fluctuations in the observed hazards in the tail ends of follow-up" hence use of a constant hazards may still be appropriate when length of follow-up is short (as in the Escudier 2009 study in the base case) and when sample size is small (in the Intermediate/poor subgroup). Hence, even when there was evidence the proportional hazards assumption was violated, the company preferred the constant hazards model over the time-varying fractional polynomials NMA.

The ERG assessed the data and agrees with the company's conclusions (Table 8). There was no strong evidence to doubt the proportional hazard assumption for OS (base case) and PFS (intermediate/poor subgroup).

3.1.7.6 Statistical NMA approaches used – constant hazards

The constant hazards NMA was performed using a regression model with a contrast-based normal likelihood for the log HR (and corresponding standard error) of each trial (or comparison) in the network (cited according to 'Dias et al' with no reference provided. The ERG believe this refers to NICE Decision Support Unit Technical Support Document (TSD) 2).[25]

Normal non-informative prior distributions for the parameters were estimated with a mean of 0 and a variance of 10,000.

As there were no closed loops in the networks (i.e. there were no direct and indirect comparisons of the same treatments), an evaluation of network internal consistency was not required.

Table 8 The company and ERG interpretation of proportional hazards assumptions

	Base case	Intermediate/poor risk
Outcome	**Studies included in NMA**	
OS	KEYNOTE, COMPARZ	KEYNOTE, CABOSUN
PFS	KEYNOTE-426, TIVO-1, COMPARZ, Motzer 2007, Escudier 2009	KEYNOTE-426, CABOSUN
The company's interpretation		
OS	PH assumption not violated. Pembrolizumab+axitinib vs sunitinib and pazopanib vs sunitinib did not vary significantly over time (CS B.2.9.4)	PH assumption violated (KEYNOTE-426). However, given low numbers of events constant HR is preferred as it is more stable (CS B.2.9.4)
PFS	PH assumption violated (CS B.2.9.4)	PH assumption not violated. HRs did not vary over time significantly (KEYNOTE-426, CABOSUN) (CS B.2.9.4)
The ERG's interpretation		
OS	KM plots unclear whether PH assumption violated for KEYNOTE-426. Log cumulative hazard plots in CS Figure 20 suggests PH assumption does not hold for KEYNOTE. However, it is unclear if these figures refer to the base case or intermediate/poor subgroup. Unclear whether PH assumption violated in TIVO-1 and COMPARZ	PH assumption violated in CABOSUN, unclear in KEYNOTE-426
PFS	KM plots unclear whether PH assumption violated for KEYNOTE-426. Log cumulative hazard plots in CS Figure 27 suggests PH assumption does not hold for KEYNOTE. However, it is unclear if these figures refer to the base case or intermediate/poor subgroup. PH assumption violated in TIVO-1, COMPARZ, Escudier 2009. Unclear whether PH assumption violated in Motzer 2007.	PH not violated in CABOSUN (NICE TA542 presents Schoenfeld residuals and log cumulative hazard plots). Unclear whether PH assumption violated in KEYNOTE-426.

PH = Proportional hazard; KM = Kaplan Meier

The ERG replicated the company's constant hazard NMA to check consistency in results using the TSD 2 code from for contrast data.[25] When using the company's data reported in CS Tables 22,24,26 & 28 results were comparable. However, when the ERG examined the underlying data from the published RCTs a number of discrepancies were found in the company's data and calculations. These are shown in Appendix 9.33 of this report. The results showed differences in PFS for tivozanib (base case) and cabozantinib (intermediate/poor risk subgroup). However, as the ERG's analyses led to slightly higher hazard ratios for these treatments the CS scenario can be viewed as conservative.

3.1.7.7 Statistical NMA approaches used – fractional polynomials

The CS cites fractional polynomial methodology introduced by Jansen[26] (CS Appendix D.1.2.3). 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.

Two orders of fractional polynomial 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 p1=0 would be equivalent to a Weibull model, and a first order model with p1=1 would correspond to a Gompertz model. Survival distributions were considered using the multivariate NMA framework: Weibull, Gompertz, and second order fractional polynomials including p_1=0 or 1 and p_2= -1, 0.5, 0, 0.5, or 1. The ERG asked the company to clarify the rationale for testing fractional polynomials with a relatively narrow range of powers (i.e. p_1 in the range 0,1, and p_2 in the range -1 to +1) (clarification question A11). The company responded that in their experience fractional polynomial models with p2=-2 or 2 were too sensitive to outliers and therefore did not reflect underlying hazard rates.

Model fitting

The deviance information criterion (DIC) was used to compare the goodness-of-fit of the competing fractional polynomial survival models. The model with the lowest DIC was chosen for analysis.

The ERG asked the company to clarify whether the choice of model was influenced by other considerations, such as the clinical plausibility of the model chosen with respect to PFS and OS curves as observed in the trials (clarification question A13). The company responded that clinical plausibility was examined by cross-referencing time-varying HRs against constant HRs from published studies and checking if results were stable across fractional polynomial models. However, the company do not elaborate further on this process and whether/how it informed their choice of model.

The DIC model fit estimates for the NMA are provided in CS Appendix M. However, only the p1=0,1 and p2=0,1 model fit results are presented. In clarification question A12, the ERG asked the company to present the model fit statistics for all the fractional polynomial models considered, to permit independent assessment of all the DIC values. The company provided a detailed appendix to the clarification responses including time-varying hazard plots and parameterisations of all fitted 1st and 2nd order fractional polynomials.

The best fitting fractional polynomial models were:

- Base case PFS – 2nd order FP model with p1=0, p2=0.
- Base case OS – 2nd order FP model with p1=0, p2=0.
- Intermediate/poor risk subgroup PFS - 2nd order FP model with p1=0, p2=0.
- Intermediate/poor risk subgroup OS – 2nd order FP model with p1=1, p2=0.

NMA results are presented for the best fitting fractional polynomials in CS Appendix M (August 2018 KEYNOTE-426 data cut) and Appendix N (January 2019 KEYNOTE-426 data cut). Results are presented as time-varying hazard ratio plots; tabulated time-varying hazard ratios (at three month intervals up to 18 months); and basic parameter estimates (d0/ d1 estimate and variance; correlation). The ERG asked the company to provide the results for each of the fractional polynomial models fitted (i.e. first order and second order p_1=0 or 1 and p_2= -1, 0.5, 0, 0.5, or 1), to enable comparison of the variation in hazard ratios between different order models

(clarification questions A12 and A14). The company provided these in the appendix to the clarification response document.

Given that the appraisal committee in NICE TA512[27] raised concerns that choice of fractional polynomial model had a substantive impact on cost-effectiveness and thereby uncertainty, we examined the impact of alternative fractional polynomial models with similar fit (Table 9) (see ERG scenario analyses in section 4.4 of this report for the cost effectiveness results). For PFS in the base case, the ERG selected the next best fitting based on DIC values (2nd order FP p1=1, p2=0). For OS in the intermediate/poor risk subgroup, since the next best fitting model (2nd order FP p1=1 p2=1) had a very similar fit to the company's best fitting model ([redacted]) for our scenario we chose the model with the third lowest DIC (2nd order FP p1=0, p2=1). We considered that the results of this fractional polynomial model ([redacted]).
[redacted] were more aligned to the respective Kaplan Meier OS curves from the KEYNOTE-426 and CABOSUN trials, and thus in our view, are more clinically plausible. (NB. all three fractional polynomial models showed no appreciable difference in fit, commonly interpreted as a difference in DIC of 2-3 or less).

The parameters of the different models were estimated using a Markov Chain Monte Carlo (MCMC) method implemented in the OpenBUGS software package. A first series of iterations from the OpenBUGS sampler were discarded as 'burn-in', and the inferences were based on additional iterations using two chains. All analyses were performed using R version 3.0.3 (http://www.r-project.org/) and OpenBUGS version 3.2.3 (OpenBUGS Project Management Group). In response to ERG clarification question A17, the company provided R code and OpenBUGS code for the models. However, the R code provided is incomplete and doesn't state which packages are used or defined the functions used. Furthermore, the data provided in the numerous spreadsheets referring to each study is not the exact data used in the NMA. Instead, it is presented as probabilities of death, the interval is unclear, as is how/whether they use the numbers at risk. The format of the data needed for the code is also unclear and initial values are not provided. Nevertheless, the ERG was able to validate the OpenBUGS code provided against that provided in published papers and is satisfied it has been conducted correctly.

Table 9 Company selected fractional polynomial model and ERG scenarios

	OS	PFS
Base case		
Company FP choice	PH assumption not violated – constant hazards used	2nd order FP p1=0, p2=0. [redacted] (best fitting model)
ERG scenario	PH assumption not violated – constant hazards used	2nd order FP p1=1, p2=0. Clarification responses appendix Tables 43, 44 [redacted] (second best fitting model)
Intermediate/poor risk subgroup		
Company FP choice	2nd order FP p1=1, p2=0 [redacted] (best fitting model)	PH assumption not violated – constant hazards used
ERG scenario	2nd order FP p1=0, p2=1 Clarification responses appendix Tables 129, 130. [redacted] (third best fitting model)	PH assumption not violated – constant hazards used

FP= Fractional polynomial; DIC = Deviance information criterion

ERG conclusion

Based on the information provided the ERG considers that the methods used to implement the two NMA methods are appropriate and correspond to the methods specified in the original methodological texts.

3.1.7.8 Choice between random effects and fixed effect models

Fixed effects models were chosen for all the NMAs presented. Random effects models are preferred in networks such as this which include small numbers of trials and where there is the potential for clinical heterogeneity[28]. However, the CS states that insufficient trials were available to achieve stable estimates of between-study heterogeneity. The ERG asked the company if an informative prior could have been used for random effects as specified by Turner et al[29] and if so to re-run the NMAs using a random effects model (clarification question A10). The company responded that collection and validation of meta-epidemiologic data as proposed by Zondervan-Zwijnenburg (2017)[30] would not have been possible within the time frame. The company therefore did not run random effects models using published informative priors. Whilst the mean effect sizes would have been the same the use of an informative prior would have

widened the credible intervals. On balance, however, we consider the trials to be generally similar in patient and other characteristics (see 3.1.7.3 above) so a fixed effect analysis is acceptable.

3.1.7.9 Summary of ERG critique of the NMA

The CS reports two types of Bayesian approaches for indirect comparison of pembrolizumab plus axitinib with other treatments:

- NMA assuming constant hazards
- NMA assuming time-varying hazards based on fractional polynomials.

These NMAs were reported for OS and PFS outcomes. The NMA assuming constant hazards appears to be the 'primary' indirect comparison method reported in the CS.

The networks are presented as a base case analysis, which included all patients irrespective of baseline RCC risk status, and subgroup analyses for patients at intermediate/poor RCC risk and patients at favourable RCC risk. The base case includes six RCTs evaluating four of the five treatments relevant to the decision problem (pembrolizumab plus axitinib, sunitinib, pazopanib, and tivozanib). The fifth treatment, cabozantinib, is only included in the intermediate/poor risk subgroup (see below).

The ERG agrees with the decision to conduct a separate NMA for the intermediate/poor RCC risk group, as inclusion in the CABOSUN trial was restricted to patients in these risk groups, and cabozantinib is recommended by NICE only for patients at intermediate/poor risk (as defined by the IMDC criteria). Expert clinical opinion is that risk status is a key prognostic variable in RCC, and there is some evidence to suggest it is an effect modifier. Caution is required in the interpretation of the subgroup NMA analyses as they are based on a subset of the KEYNOTE-426 randomised trial population, which can increase uncertainty about the precision of treatment effects.

The NMA does not inform the economic model for the base case analysis (all patients irrespective of baseline RCC risk status). However, the NMA informs the economic model for the intermediate/poor risk subgroup analysis comparing pembrolizumab plus axitinib versus cabozantinib.

In terms of clinical heterogeneity, the trials included in the NMA can be considered similar to each other in terms of patient demographic and prognostic characteristics and in sunitinib treatment regimens. An exception to this is the CABOSUN trial which differed from the other trials on a number of characteristics (e.g. smaller sample size; only included patients at intermediate or poor RCC risk). These differences are likely to be related to the intermediate/poor risk status of the patients in this trial. The impact of these differences on the results of the NMA are mitigated by exclusion of CABOSUN in the base case NMA, and its inclusion in the intermediate/poor RCC risk subgroup. Overall, the trials were considered to have low risk of bias, aside from bias associated with open-label trials.

The *a priori* rationale for using both constant hazards and time-varying hazards NMA assumptions in the CS is not explicitly stated. The company discusses these assumptions based on a comparison of the results of the constant hazards and the time-varying hazards fractional polynomial NMAs. The company concluded proportional hazards did not hold for PFS (base case) and OS (intermediate/poor subgroup). However, the company maintained that despite the violation of the proportional hazards assumption, the use of constant hazards is more appropriate than time-varying fractional polynomials when length of follow-up is short, or sample size is small. The ERG assessed the data and agrees with the company's conclusions.

The constant hazards NMA was conducted according to standard methods as recommended by the NICE DSU.[25] The fractional polynomials model was conducted according to methods proposed in a journal paper by author Jansen.[26] The DIC was used to compare the goodness-of-fit of the competing fractional polynomial survival models. The model with the lowest DIC was chosen for analysis.

Given that the appraisal committee in NICE TA512 (tivozanib)[27] raised concerns that the choice of fractional polynomial model had a substantive impact on cost-effectiveness and thereby uncertainty, we examined the impact of alternative fractional polynomial models with similar fit in an ERG scenario analysis (see section 4.4 of this report).

Fixed effect models were chosen by the company for all the NMAs presented. Random effects models are preferred in networks such as this which include small numbers of trials and where there is potential for heterogeneity. The CS states that insufficient trials were available to achieve stable estimates of between-study heterogeneity, and it was not possible to use an

informative prior for a random effects analysis. The ERG concurs that fixed effect model is acceptable given the general low clinical heterogeneity (see above).

Overall, the ERG considers the methods and assumptions used to conduct the NMAs to have been appropriately exercised, though with some uncertainties due to relatively small data sets.

3.2 Summary statement of company's approach to evidence synthesis

The ERG's appraisal of the CS systematic review of clinical effectiveness is summarised in Table 10.

Table 10 ERG critical appraisal of company's systematic review of clinical effectiveness

Item	ERG response
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Yes. Although the eligibility criteria used for population includes a narrower population of RCC with clear cell component.
2. Is there evidence of a substantial effort to search for all relevant research?	Yes. There was a sufficient effort to search for relevant research. Although the search was around five months out-of-date the ERG has run targeted searches for recent evidence and no studies of relevance appear to have been missed.
3. Is the validity of included studies adequately assessed?	Yes. Although the ERG differed with some of the company's judgements (more favourably).
4. Is sufficient detail of the individual studies presented?	Yes. Sufficient details were reported.
5. Are the primary studies summarised appropriately?	Yes. The CS summaries of the key characteristics of the relevant trials and their results appeared accurate and appropriate.

The ERG considers the systematic review processes undertaken by the company were reasonable (two reviewers undertook all stages) with the exception of post hoc inclusion of two trials in the NMA. The evidence presented reflects the decision problem with the exception of the population having a more precise definition for clear cell (+/- sarcomatoid features) and

some participants may have been treated at an advanced stage previously. There is a low chance of systematic error in the results of the systematic review.

3.3 Summary of submitted evidence

The ERG has summarised results from the August 2018 data-cut and noted similarities or differences from the January 2019 data-cut. We consider BICR assessments of response to be the least biased, however investigator assessments are also noted for comparison.

3.3.1 Survival

At the August 2018 data-cut and a median follow-up of 13.2 months (range 0.1 to 21.5 months) in the pembrolizumab plus axitinib arm and 12.1 months (range 0.4 to 22.0 months) in the sunitinib arm, median OS was not reached in either group. Overall survival rates at 6, 12 and 18 months favoured pembrolizumab plus axitinib (Table 11). The HR for OS was 0.53 (95% CI 0.38, 0.74), p=0.00005. Results from the January 2019 data cut [redacted], CS Appendix O Table 6.

Median PFS based on BICR was 15.1 months with pembrolizumab plus axitinib and 11.1 months with sunitinib, HR 0.69 (95% CI 0.57 to 0.84, p=0.00014) at the August 2018 data-cut. PFS rates at 6, 12 and 18 months favoured pembrolizumab plus axitinib (Table 11), although 95% confidence intervals overlapped at 6 and 18 months. Results from the January 2019 data cut [redacted] (CS Appendix O Table 7). PFS at the August 2018 data-cut based on investigator assessment was less favourable but remained statistically significant (HR 0.82, 95% CI 0.67 to 1.00, p=0.022) (CS Appendix L Table 3).

Although both OS and PFS inform the economic model, hazard rates of PFS were only based on the observed Kaplan-Meier curve up to week 13, with parametric models fitted to data after this time point (see section 4.3.5).

3.3.2 Response rates

At the August 2018 data-cut, the ORR was 59.3% in the pembrolizumab plus axitinib arm and 35.7% in the sunitinib arm based on BICR according to RECIST 1.1, a difference of 23.6% (95% CI 17.2 to 29.9, p<0.0001).

Table 11 Survival outcomes at August 2018 data-cut

	Pembrolizumab + axitinib (n=432)	Sunitinib (n=429)	Treatment effect (95% CI), p value
Overall survival			
Median OS, months	Not reached	Not reached	HR 0.53 (0.38, 0.74), p=0.00005
6 month OS rate, % (95% CI)	94.9 (92.3, 96.6)	89.0 (85.6, 91.6)	NR
12 month OS rate, % (95% CI)	89.9 (86.4, 92.4)	78.3 (73.8, 82.1)	NR
18 month OS rate, % (95% CI)	82.3 (77.2, 86.3)	72.1 (66.3, 77.0)	NR
PFS			
Median PFS, months	15.1 (12.6, 17.7)	11.1 (8.7, 12.5)	HR 0.69 (0.57, 0.84), 0.00014
6 month PFS rate, % (95% CI)	74.0 (69.5, 77.9)	66.0 (61.1, 70.4)	NR
12 month PFS rate, % (95% CI)	59.6 (54.3, 64.5)	46.2 (40.6, 51.6)	NR
18 month PFS rate, % (95% CI)	41.1 (33.5, 48.5)	32.9 (25.4, 40.5)	NR

NR, not reported.

A complete response was experienced by 5.8% of the pembrolizumab plus axitinib arm and 1.9% of the sunitinib arm (Table 12).

[redacted].

The difference in DCR based on BICR was 11% (95% CI 4.8 to 17.0) in favour of pembrolizumab plus axitinib, with [redacted] at the January 2019 data-cut (CS Appendix O Table 9). The difference based on investigator assessments was lower at 6.6% (95% CI 4.8 to 17.0) at August 2018 (CS Appendix L Table 10).

Median DOR based on BICR in people with a CR or PR was not reached at the August 2018 data-cut in the pembrolizumab plus axitinib arm and was 15.2 months in the sunitinib arm (Table 12). Median DOR based on investigator assessment was 18.0 months (range 1.3+ to 18.2+) and 15.2 months (range 1.2+ to 15.4+), respectively (CS Appendix L Table 7). At the January 2019 data-cut these outcomes were [redacted](CS Appendix O Table 2).

Table 12 Response rates and DOR based on BICR at August 2018 data-cut

	Pembrolizumab + axitinib (n=432)		Sunitinib (n=429)		Difference (95% CI), p value
	n (%)	95% CI	n (%)	95% CI	
Objective response rate (CR+PR)	256 (59.3)	54.5, 63.9	153 (35.7)	31.1, 40.4	23.6 (17.2, 29.9), p<0.0001
Disease control rate (CR+PR+SD ≥ 6 months)	309 (71.5)	67.0, 75.7	260 (60.6)	55.8, 65.3	11.0 (4.8, 17.0) p=NR
CR	25 (5.8)	3.8, 8.4	8 (1.9)	0.8, 3.6	NR
PR	231 (53.5)	48.6, 58.3	145 (33.8)	29.3, 38.5	NR
SD	106 (24.5)	20.5, 28.9	169 (39.4)	34.7, 44.2	NR
PD	47 (10.9)	8.1, 14.2	73 (17.0)	13.6, 20.9	NR
Non-evaluable[a]	8 (1.9)	0.8, 3.6	6 (1.4)	0.5, 3.0	NR
No assessment[b]	15 (3.5)	2.0, 5.7	28 (6.5)	4.4, 9.3	NR
Median DOR (in CR or PR), months (range)	Not reached (1.4+ to 8.2+)		15.2 (1.1+ to 15.4+)		NR

[a] post-baseline assessment(s) available however not being evaluable (i.e., all post-baseline assessment(s) with insufficient data for assessment of response per RECIST 1.1. or CR/PR/SD < 6 weeks from randomisation). [b] no post-baseline assessment available for response evaluation. For best overall response of CR and PR, only confirmed responses are included. CR, complete response; DOR: duration of response; NR, not reported; PD, progressive disease; PR, partial response; SD, stable disease. '+' indicates there was no progressive disease by the time of last disease assessment.

3.3.3 Health related quality of life

Utilities at baseline and end of study measured with the EQ-5D-3L were not presented in the CS. These were provided by the company in response to clarification request A8, and changes from baseline are summarised in Table 13. There was no statistically significant difference between treatments.

The CS states there were no clinically meaningful differences in EQ-5D VAS between baseline and week 30 in either group, and changes from baseline at week 30 were similar between groups (Table 13). The CS does not define what a clinically meaningful difference is, however the ERG notes that a minimal important difference of 7 has been applied in kidney cancer previously.[31] [redacted] were found at the January 2019 data-cut (CS Appendix O Table 10). Similarly, no differences between groups were found in EORTC QLQ-C30 global health status/QoL score (Table 13). Results from the January 2019 data-cut were not reported.

Selected functional and symptom scales of the EORTC-QLQ-30 were presented in CS Appendix L Figure 6. The results for all EORTC-QLQ-30 scales were provided by the company in response to clarification request A7 in Figures A7.1 and A7.2. A greater worsening of the

diarrhoea symptom scale was observed in the pembrolizumab + axitinib group. Reporting of diarrhoea as an adverse event is discussed in section 3.3.6. The other scales (Global health status/QoL, functional scales, symptom scales and items) were similar between treatments.

Table 13 Patient reported outcomes at August 2018 data-cut

	Pembrolizumab + axitinib (n=432)	Sunitinib (n=429)	Difference in LS mean (95% CI), p value
EQ-5D-3L index			
Change from baseline, LS mean (95% CI)	n=427 -0.005 (-0.026, 0.016)	n=421 -0.013 (-0.035, 0.010)	0.007 (-0.022, 0.037) p=0.619
EQ-5D VAS			
Change from baseline, LS mean (95% CI)	n=427 -3.31 (-5.18, -1.43)	n=421 -1.92 (-3.90, 0.05)	-1.38 (-3.89, 1.12), p=0.277
EORTC QLQ-C30 global health status/QoL score			
Change from baseline, LS mean (95% CI)	n=427 -4.05 (-6.03, -2.07)	n=423 -2.35 (-4.44, -0.26)	-1.70 (-4.34, 0.94), p=0.207

Source: CS Table 19; CS Appendix L Table 13; clarification response Table A8.

3.3.4 Sub-group analyses for overall survival and PFS

As described earlier in section 3.1.6, a number of patient subgroups were analysed in the KEYNOTE-426 trial. At the August 2018 data cut, OS results for the subgroups were consistent with the overall effect (subgroup HRs ranging from 0.2 to 0.69 with wider confidence intervals for some; overall HR 0.53), (CS Appendix E Figure 1).

[redacted]

PFS was consistent across all subgroups, with HRs ranging from 0.54 to 0.87 (overall HR 0.69), (CS Appendix E Figure 2).

[redacted].

No statistical tests of interaction were presented and the CS states these subgroups lack statistical power, therefore these results should be viewed with caution.

ERG conclusion

OS and PFS results were in favour of pembrolizumab plus axitinib over sunitinib, although median OS was not reached in either group. Objective response rate and disease control rate were also in favour of pembrolizumab plus axitinib. There were no differences between treatments for the EQ-5D-3L index, EQ-5D VAS or EORTC QLQ-C30 global health status/QoL score and most other scales of the EORTC QLQ-C30, apart from a greater worsening on the diarrhoea scale with pembrolizumab plus axitinib. Subgroup analyses of OS and PFS were consistent with the overall effect [redacted].

3.3.5 Network meta-analysis results

For brevity summarise the results of the constant hazards NMA, with brief reference to the results of the time-varying fractional polynomial NMA. Please refer to section 3.1.7 for a description of the evidence networks for the base case and subgroup analyses, and the statistical procedures used to conduct the NMAs.

3.3.5.1 Progression free survival

Table 14 reports the NMA inputs to the constant HRs for the outcome of PFS in the base case.

Table 14 Constant HRs NMA (fixed effects) for PFS, NMA base case

Trial	Reference	Intervention	HR Int. vs. Ref	Log HR (SE) Int. vs. Ref
COMPARZ	Sunitinib	Pazopanib	[redacted]	[redacted]
Escudier 2009	IFN-α	Sorafenib	[redacted]	[redacted]
KEYNOTE-426 (IA1 Aug 2018 data cut)	Sunitinib	Pembrolizumab + axitinib	[redacted]	[redacted]
Motzer 2007	Sunitinib	IFN-α	[redacted]	[redacted]
TIVO-1*	Sorafenib	Tivozanib	[redacted]	[redacted]

Note: * denotes trials in grey used subgroup first-line data
Grey rows represent treatments that were not of interest but facilitated indirect treatment comparisons for treatments of interest.
Reproduced from CS Table 22.

The CS reports that treatment with pembrolizumab plus axitinib resulted in a [redacted] in the duration of PFS compared to all relevant competing interventions including [redacted] (Table 15).

Table 15 HRs estimated from fixed-effects constant hazard NMA of PFS; base case

Sunitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	**IFN-a**	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	**Sorafenib**	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	**Pazopanib**	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	**Tivozanib**	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Pembrolizumab +axitinib**

Note: Each cell represents the comparison (hazard ratio and 95% CrI) of the row treatment versus the column treatment.
Grey cells represent treatments that were not of interest but facilitated indirect treatment comparisons for treatments of interest.
All bolded values are statistically meaningful at the 0.05 significance level.
Reproduced from CS Table 23.

Results using the January 2019 KEYNOTE-426 data-cut show [redacted] results.

The results of the time-varying fractional polynomial NMA showed that pembrolizumab plus axitinib [redacted] (relative to sunitinib) than tivozanib and pazopanib, and achieved [redacted] compared to the other comparators over time.

Table 16 reports the NMA inputs to the constant HRs for the outcome of PFS in the intermediate/poor risk subgroup.

Table 16 Error! Reference source not found.**Constant HRs for PFS; intermediate/poor risk subgroup**

Study	Reference	Intervention	HR Int. vs. Ref	Log HR (SE) Int. vs. Ref
CABOSUN	Sunitinib	Cabozantinib	[redacted]	[redacted]
KEYNOTE-426 (IA1 - Aug 2018 data-cut)	Sunitinib	Pembrolizumab + axitinib	[redacted]	[redacted]

Reproduced from CS Table 26

The CS reports that both cabozantinib and pembrolizumab plus axitinib were associated with lower HRs compared to sunitinib, indicating longer PFS (Table 17).

Table 17 HRs estimated from fixed-effects constant hazard NMA of PFS; intermediate/poor risk subgroup

Sunitinib	[redacted]	[redacted]
[redacted]	**Cabozantinib**	[redacted]
[redacted]	[redacted]	**Pembrolizumab + axitinib**

Note: Each cell represents the comparison (hazard ratio and 95% CrI) of the row treatment versus the column treatment.
All bolded values are statistically meaningful at the 0.05 significance level.
Reproduced from CS Table 27
[redacted] were obtained when using the January 2019 data cut.

The results of the time-varying fractional polynomial NMA showed that pembrolizumab plus axitinib had a [redacted] compared to sunitinib up to six months and cabozantinib had a [redacted] compared to sunitinib [redacted]. There was [redacted] between pembrolizumab plus axitinib and cabozantinib [redacted]

3.3.5.2 Overall survival

Table 18 reports NMA inputs to the constant HRs for the outcome of OS in the base case.

Table 18 Constant HRs for OS; NMA base case

Study	Reference	Intervention	HR Int. vs. Ref	Log HR (SE) Int. vs. Ref
COMPARZ	Sunitinib	Pazopanib	[redacted]	[redacted]
KEYNOTE-426 (IA1 - Aug 2018 data-cut)	Sunitinib	Pembrolizumab + axitinib	[redacted]	[redacted]

Reproduced from CS Table 24

As can be seen, tivozanib is omitted from this network. The CS reports that treatment with pembrolizumab plus axitinib resulted in a [redacted] in the duration of OS compared to pazopanib ([redacted]) and sunitinib ([redacted]) (Table 19).

Table 19 HRs estimated from fixed-effects constant hazard NMA of OS; base case

Sunitinib	[redacted]	[redacted]
[redacted]	**Pazopanib**	**1.73 (1.21, 2.49)**
[redacted]	[redacted]	**Pembrolizumab + axitinib**

Note: Each cell represents the comparison (hazard ratio and 95% CrI) of the row treatment versus the column treatment. All bolded values are statistically meaningful at the 0.05 significance level.

Reproduced from CS Table 25

[redacted] were obtained when using the January 2019 data cut.

The results of the time-varying fractional polynomial NMA showed that pembrolizumab plus axitinib versus sunitinib was associated with [redacted] HRs over time compared to pazopanib versus sunitinib. There [redacted] between treatments [redacted]

Table 20 reports the NMA inputs to the constant HRs for the outcome of OS, based on the intermediate/poor risk subgroup.

Table 20 Constant HRs for OS; intermediate/poor risk subgroup

Study	Reference	Intervention	HR Int. vs. Ref	Log HR (SE) Int. vs. Ref
CABOSUN	Sunitinib	Cabozantinib	[redacted]	[redacted]
KEYNOTE-426 (IA1 - Aug 2018 data-cut)	Sunitinib	Pembrolizumab + axitinib	[redacted]	[redacted]

Reproduced from CS Table 28

The CS reports that pembrolizumab plus axitinib resulted in a [redacted] in OS compared to sunitinib ([redacted]) (Table 21).

Table 21 HRs estimated from fixed-effects constant hazard NMA of OS; intermediate/poor risk subgroup

Sunitinib	[redacted]	[redacted]
[redacted]	**Cabozantinib**	[redacted]
[redacted]	[redacted]	**Pembrolizumab + axitinib**

Note: Each cell represents the comparison (hazard ratio and 95% CrI) of the row treatment versus the column treatment. All bolded values are statistically meaningful at the 0.05 significance level.
Reproduced from CS Table 29

[redacted] were obtained when using the January 2019 data cut from KEYNOTE-426.

The results of the time-varying fractional polynomial NMA showed that pembrolizumab plus axitinib was associated with [redacted] compared to cabozantinib versus sunitinib. However, differences between pembrolizumab plus axitinib and cabozantinib [redacted] due to [redacted]

As stated in section 3.1.7, the ERG asked the company to run separate NMA scenario analyses for (i) intermediate risk patients, and (ii) poor risk patients (clarification question A21). The results of the scenario analyses (constant hazards NMA only) showed that HRs were [redacted]. Due to the relatively small sample sizes in the poor risk subgroups it cannot necessarily be concluded that there are significant differences in effect (PFS) between intermediate and poor risk subgroups.

3.3.6 Summary of adverse events

Safety data for the KEYNOTE-426 trial are reported in CS section B.2.10 (August 2018 data cut) and also in CS Appendices F (additional data from August 2018) and O (January 2019 data cut). A summary overview of all AEs is presented in CS Table 33 and reproduced in Table 22. Data on discontinuations and deaths due to AEs were presented in CS Appendix O p211-212. Where possible the ERG has cross-checked these data with the KEYNOTE-426 journal publication or the CSR. As reported in the CS, the overall incidence of AEs was generally similar in the pembrolizumab plus axitinib and sunitinib trial arms. The majority of patients in both treatment arms reported at least one any Grade AE (pembrolizumab plus axitinib 98.4%;

sunitinib 99.5%). In the pembrolizumab plus axitinib arm drug-related AEs were experienced by 96.3% of patients; 75.8% any Grade 3-5 AE and 7.7% patients discontinued both drugs simultaneously owing to an AE. In the sunitinib arm 97.6% of patients experienced a drug-related AE; 70.6% any Grade 3-5 AE and 13.9% discontinued sunitinib owing to an AE. Rates of discontinuations of either pembrolizumab or axitinib or both are reported in CS Table 33.

The rate of serious adverse events (SAEs) was higher in the pembrolizumab plus axitinib group; 40.3% of participants reported SAEs in the pembrolizumab plus axitinib arm compared with 31.3% in the sunitinib arm (CS Appendix F Table 2 provides details of specific SAEs the most common in the pembrolizumab plus axitinib arm being diarrhoea, 2.8%). Deaths due to AEs occurred in 2.6% of the pembrolizumab plus axitinib arm (four of these were drug-related AEs: myasthenia gravis, myocarditis, necrotizing fasciitis, and pneumonitis) and 3.5% of the sunitinib arm (7 were drug-related AEs). Adverse event rates were [redacted] at the January 2019 data cut (Table 22) with the exception of [redacted] (although the difference between trial arms was similar to the 2018 data-cut).

As reported in the CS (Table 30) there were differences in treatment exposure between the two arms of the KEYNOTE-426 trial. The mean duration on any therapy was 320.6 days (total exposure 4766.94 person-months) in the pembrolizumab plus axitinib arm and in the sunitinib arm this was 255.6 days (total exposure 3924.64 person-months) (Table 23). The CS presents exposure-adjusted event rates for the key adverse events to account for the different times on therapy. This is presented as the event rate per 100 person-months of exposure, calculated as the event count multiplied by 100 divided by person-months of total exposure of all participants in that group (time between the first dose date plus 1 day and the earlier of the last dose date plus 30 or the database cut-off date). No further details were presented in the CS and the ERG was unable to find details in the CSR or trial protocol. Note that the adjusted rate is based on the count of events, rather than the number of people experiencing the event, therefore includes multiple occurrences of events. Details of any censoring in the arms was not reported. The exposure-adjusted incidence rate is most appropriate when the risk of each event is constant over the duration of follow-up,[32] but the ERG notes some adverse events may be more likely to occur earlier or later with treatment (see below).

When the exposure-adjusted rates are considered, the CS states that there were no clinically meaningful differences in overall rates including SAEs (CS p66), CS Table 34 (reproduced in Table 24).

Table 22 Summary of adverse events in KEYNOTE-426, All Subjects as Treated (ASaT) population

Event, %	Pembrolizumab plus axitinib, n=429, 2018 data cut ([redacted])	Sunitinib, n=425, 2018 data cut ([redacted])
Any AE	98.4 ([redacted])	99.5 ([redacted])
Any drug-related AE	96.3 ([redacted])	97.6 ([redacted])
Grade 3-5 AE	75.8 ([redacted])	70.6 ([redacted])
Grade 3-5 drug-related AE	62.9 ([redacted])	58.1 ([redacted])
SAE	40.3 ([redacted])	31.3 ([redacted])
Treatment discontinuation for AE	7.7[a] ([redacted])	13.9[b] ([redacted])
Drug-related AE leading to discontinuation	6.3[c] ([redacted])	10.1 ([redacted])
Death related to AE	2.6 ([redacted])	3.5 ([redacted])

[a]discontinuation of both drugs simultaneously; [b]CS Appendix F reports 13.4% [c]discontinuation of both drugs

Table 23 Overview of duration on any therapy in KEYNOTE-426, ASaT population

Duration[a], days	Pembrolizumab plus axitinib, n=429	Sunitinib, n=425
Mean (SD)	320.6 (163.2)	255.6 (165.6)
Median (range)	317 (1 to 646)	238 (2 to 623)

[a]days between first dose date and last dose date

Table 24 Exposure adjusted summary of AE in KEYNOTE-426, ASaT population

	Pembrolizumab plus axitinib, n=429	Sunitinib, n=425
Total exposure[a] in person-months	4766.94	3924.64
Rate (event count / 100 person-months)[b]		
Any AE	147.20	179.69
Any drug-related AE	83.74	126.25
Grade 3-5 AE	17.75	20.97
Grade 3-5 drug-related AE	11.56	14.40
SAE	5.96	5.12
AE leading to discontinuation	3.78	1.66
Drug-related AE leading to discontinuation	3.19	1.20
Death related to AE	0.23	0.41

[a] defined as the time between the first dose date plus 1 day and the earlier of the last dose date plus 30 or the database cut-off date. [b] event rate per 100 person-months of exposure = event count *100/person-months of exposure.

3.3.6.1 Commonly reported AEs

The most common types of AEs (any grade) and drug-related AEs (any grade) can be seen in Table 25. Participants receiving pembrolizumab plus axitinib had a greater likelihood of

dysphonia, arthralgia, diarrhoea and pruritis amongst others. Participants receiving sunitinib were more likely to have anaemia, thrombocytopenia, dysgeusia, and neutropenia. CS Figure 15 displays between treatment comparisons of the most common AEs sorted by risk difference. The frequency of drug-related AEs showed similar patterns (Table 25) and the drug-related AEs reported by the later data cut (January 2019) were very similar (CS Appendix O, Table 14).

Table 25 Commonly reported AEs and drug-related AEs in the KEYNOTE-426 trial, ASaT population, August 2018 data-cut

%	**Any AE (≥15% in at least one arm)**		**Drug-related AE**	
	Pembrolizumab plus axitinib n=429	**Sunitinib n=425**	**Pembrolizumab plus axitinib n=429**	**Sunitinib n=425**
Diarrhoea	54.3	44.9	49.0	41.2
Hypertension	44.5	45.4	41.7	43.3
Fatigue	38.5	37.9	30.3	33.4
Hypothyroidism	35.4	31.5	31.5	28.0
Decreased appetite	29.6	29.4	21.9	24.9
Palmar-plantar erythrodysaesthesia syndrome	28.0	40.0	27.7	39.5
Nausea	27.7	31.5	21.2	26.1
ALT increased	26.8	15.1	23.8	12.7
AST increased	26.1	16.2	22.6	13.9
Dysphonia	25.4	3.3	22.8	2.8
Cough	21.2	13.6	7.5	2.8
Constipation	20.7	14.6	7.2	6.8
Arthralgia	18.2	6.1	12.1	3.5
Weight decreased	17.7	11.1	9.6	8.5
Proteinuria	17.5	11.1	15.4	9.2
Dyspnoea	16.1	10.8	6.5	3.8
Headache	15.9	16.2	8.2	7.8
Stomatitis	15.6	20.9	14.2	20.2
Asthenia	15.2	14.8	11.7	12.7
Pruritus	15.2	5.9	12.4	4.2
Vomiting	15.2	18.6	7.9	13.2
Mucosal inflammation	13.3	21.9	12.8	21.2
Dysgeusia	11.0	30.8	9.3	30.4
Anaemia	7.9	23.5	2.8	16.2
Platelet count decreased	3.7	18.1	3.3	17.9
Thrombocytopenia	2.6	23.3	1.9	22.1
Neutropenia	1.9	19.3	1.4	18.6

ALT: Alanine aminotransferase; AST: Aspartate aminotransferase

While not directly comparable, the exposure-adjusted values for these AEs according to observation period (overall rates [CS Table 36], not presented for drug-related events) appear to mirror the same trends. The company notes that the exposure-adjusted total AE rate was lower

for pembrolizumab plus axitinib compared with sunitinib during months 0 to 3, similar during months 3 to 6 and higher from 6 months onwards.

The adverse events in Table 25 are summarised for any grade. Of these, the most common grade 3 to 5 AEs and grade 3 to 5 drug-related AEs (occurring in ≥5% of any arm) are shown in Table 26 (reproduced from CS Table 37 and 38). Apart from fatigue, most of these grade 3 to 5 events were drug-related. CS Appendix F Figure 2 presents between treatment comparisons for grade 3 to 5 AEs sorted by risk difference. Pembrolizumab plus axitinib had a higher risk of increased alanine aminotransferase, increased aspartate aminotransferase and diarrhoea. Sunitinib had a higher risk of fatigue, thrombocytopenia and neutropenia among others. The exposure-adjusted values for the individual AEs were not reported in the CS (only the overall event rates as discussed above, Table 24). The grade 3 to 5 AE rates were [redacted] with longer follow-up (data cut January 2019), CS Appendix O Tables 13 and 15.

Table 26 Commonly reported Grade 3 – 5 AEs and drug-related Grade 3 – 5 AEs in KEYNOTE-426, August 2018 data-cut

%	Grade 3-5 AE (≥5% in either group)		Grade 3-5 drug-related AE (≥5% in either group)	
	Pembrolizumab plus axitinib n=429	Sunitinib n=425	Pembrolizumab plus axitinib n=429	Sunitinib n=425
Hypertension	22.1	19.3	21.2	18.4
ALT increased	13.3	3.1	12.1	2.6
Diarrhoea	9.1	4.7	7.2	4.5
AST increased	7.0	2.4	6.8	1.6
Palmar-plantar erythrodysaesthesia syndrome	5.1	3.8	5.1	3.5
Fatigue	2.8	6.6	[redacted]	[redacted]
Neutropenia	0.2	6.6	0.2	6.6
Thrombocytopenia	0.0	5.9	0.0	5.2
Neutrophil count decreased	0.2	6.8	0.2	6.8
Platelet count decreased	0.2	7.3	0.2	7.3

[a]data for completeness although incidence <5%. ALT: Alanine aminotransferase; AST: Aspartate aminotransferase
Reproduced from CS Table 37 and 38

3.3.6.2 Adverse events of special interest

CS Appendix F reports that adverse events of special interest (AEOSI) are a selection of immune-related adverse events developed by the company considered during the pembrolizumab monotherapy research programme considered to be causally related to pembrolizumab. These are classified according to a medical concept which is comprised of

subcategories or preferred terms, an example given is ‘immune-related hypothyroidism’ which has preferred terms of hypothyroidism, hypothyroidic goitre, myxoedema, myxoedema coma and primary hypothyroidism. When pembrolizumab is combined with other treatments there may be overlapping adverse events and in these cases the CS says these events may then not always be immune-related. The example of hypothyroidism is given (however the CS considers hypothyroidism as an AEOSI). Similarly, the CS says that if an active control has an adverse event profile that overlaps with the preferred terms it may not be considered immune-related unless the control drug itself is an immunomodulatory agent. Clinical advice to the ERG confirms that some toxicities can be both immune-mediated or caused in other ways by other agents. Examples are diarrhoea and hypothyroidism; these can be immune-mediated but they are also side effects of TKIs via a different mechanism.

It is unclear whether these AEOSIs were specified apriori as neither the trial protocol or the CSR discuss this.

The CS reports (page 82) that the overall incidence of AEOSI was higher for pembrolizumab plus axitinib compared with sunitinib across all categories, in accordance with expectations. Additionally, there were higher rates of AEOSI in the pembrolizumab plus axitinib group than would be expected for pembrolizumab monotherapy. The higher incidence of AEOSI was mostly from thyroid-related events (hypothyroidism, hyperthyroidism, and thyroiditis). The CS also states that the majority of events were grade 1 or 2 (i.e. mild to moderate). The overall incidence of AEOSI at grades 3-5 can be seen in Table 27.

The ERG requested results of AEOSI by grade in a clarification question to the company (question A3) and the information supplied confirms that the majority of events within the preferred terms were grade 1 or 2. As none of these have incidences of >5% at grade 3 or above the ERG has not summarised these here. Reflecting the overall pattern of AEOSIs, the grade 3 AEOSIs had higher incidence in the pembrolizumab plus axitinib arm compared with the sunitinib arm.

Table 27 Grades 3-5 AEOSI by treatment group in KEYNOTE-426, August 2018 data-cut

AEOSI, %	Pembrolizumab plus axitinib, n=429	Sunitinib, n=425
Grade 3	8.4	1.6
Grade 4	1.6	0.0
Grade 5	0.7	0.2

In the CS Appendix F (page 217) it reports that there were higher incidences of hepatic AEs in the pembrolizumab plus axitinib arm (overall with one or more hepatic adverse event 40.6% pembrolizumab + axitinib; 26.6% sunitinib). Of these, hepatitis was considered an AEOSI and the incidence of hepatitis was reported to be pembrolizumab plus axitinib (2.8%) compared with sunitinib (0.5%) in CS Appendix F (page 214). In the company's response to clarification question 3 (Table A3) it can be seen that in the pembrolizumab plus axitinib arm grade 3 autoimmune hepatitis occurred in 0.5% patients; grade 4 drug-induced liver injury occurred in 0.2%; grade 3 or 4 hepatitis occurred in 1.4% and grade 3 immune-mediated hepatitis in 0.2%. These events did not occur in the sunitinib arm where there was one case of grade 5 hepatitis fulminant and one case of grade 1 hepatitis.

In summary a greater number of AEOSI were reported by the company for pembrolizumab plus axitinib, but not all of these events are necessarily immune-related (the company does not classify which) and in most cases they were not grade 3 or 4 events.

3.3.6.3 Additional sources of AE data

The ERG considered whether the adverse events reported in the pembrolizumab plus axitinib phase Ib study (KEYNOTE-035[33]}) would be informative. The phase Ib study was a dose finding study and as the dose of pembrolizumab was not directly comparable with the dose of pembrolizumab in the KEYNOTE-426 trial (2mg/kg versus 200mg respectively) we have not summarised the key AEs from the phase Ib study.

3.3.6.4 Safety overview

Overall, the CS considers that the safety profile of pembrolizumab plus axitinib is acceptable. The overall rate of AEs was similar across both arms of the trial, particularly when adjusted for exposure of the drugs. The most commonly reported AEs at grade 3 or more were hypertension, diarrhoea, alanine aminotransferase and aspartate aminotransferase increases and Palmar-plantar erythrodysaesthesia syndrome. The CS also states that the safety profile of pembrolizumab plus axitinib is generally consistent with the established safety profile of pembrolizumab monotherapy in solid tumours and the observed safety profile for axitinib monotherapy. No evidence of AEs from axitinib monotherapy was provided except for reference to published data. The ERG has checked the two references cited in the CS (CS references 31 and 47, p84) and although the AEs presented in these publications were not wholly comparable with the rates reported for the key events shown in Table 26, they appear to

show consistent effects with the KEYNOTE-426 trial. No studies of pembrolizumab as monotherapy in RCC were identified.

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 comparator therapies to pembrolizumab in treating patients with advanced RCC.
ii) a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of pembrolizumab plus axitinib is compared with sunitinib, tivozanib, pazopanib and cabozantinib for adults with untreated advanced RCC.

4.2 Company's review of published economic evaluations

A systematic search of the literature was conducted by the company to identify economic evaluations of pembrolizumab plus axitinib and comparator therapies in a patient population with unresectable advanced RCC, in addition to resource use and costs associated with treating advanced renal cell carcinoma. The following databases were searched alongside a thorough review of the grey literature: EMBASE, MEDLINE, the Cochrane Central Register of Controlled Trials (CCTR), the Cochrane Database of Systematic Reviews (CDSR), EconLit, the NHS EED and the Tufts Cost-Effectiveness Analysis Registry. An initial search was conducted between March and April 2018, followed by an updated search of all the previously searched databases. This updated search was conducted in February 2019.

The inclusion and exclusion criteria for the systematic review are listed in Table 1 of the CS Appendix G, page 221. The inclusion criteria state that primary research studies (including observational studies, RCTs and economic evaluations) and HTA documents of pembrolizumab monotherapy or in combination with another agent indicated for first-line metastatic RCC versus any comparator of interest (including placebo and best supportive care) in adults 18 years and above with unresectable metastatic RCC would be included. The exclusion criteria excluded studies of patients with early stage RCC and comparators such as surgery, radiotherapy and treatments used in adjuvant therapy.

The company's systematic literature review identified 1,351 studies and after abstract screening, 218 were deemed eligible for full-text screening. Of these, 212 studies were excluded because they did not meet the study design inclusion criteria and/or contained outcomes that were not of interest. The company identified an additional three studies through the grey literature and from the final appraisal determination of nivolumab with ipilimumab for untreated advanced renal cell carcinoma in May 2019. These studies were added to the initial six studies, bringing the total number of studies included for full review and data extraction to 10.[24, 34-41] None of these 10 studies contained a cost-effectiveness analysis for pembrolizumab in combination with axitinib. A full list of these studies is reported in Table 2 of CS Appendix G and they are all studies of comparator technologies. In the company's final inclusion, one of these 10 studies (Mickisch et al[42]) was excluded because it did not inform an HTA submission or contain a comparator relevant to this appraisal. Of the remaining nine studies, seven considered a UK NHS perspective while the remaining two were based on the perspective of the Scottish healthcare system. The results of the cost-utility analyses of these studies are reproduced below in Table 28.

Table 28 Results of cost-utility analyses for studies included in the company's search

Authors	Year	Intervention	Comparator	Incremental costs	Incremental QALYs	ICER (QALYs)
Amdahl et al	2017	Pazopanib	Sunitinib	-£912.00	0.0594	Dominant
Hoyle et al	2010	Temsirolimus	Interferon-alpha	£22,331	0.24	£94,632
Kilonzo et al	2013	Sunitinib	Pazopanib	Not reported	Not reported	£1,790
		Interferon-alpha	Pazopanib	Not reported	Not reported	£38,925
		Best supportive care	Pazopanib	Not reported	Not reported	£32,898
NICE	2009	Sunitinib	Interferon-alpha	Not reported	Not reported	£49,304
SMC	2011	Pazopanib	Sunitinib	£4,263	0.068	£62,414
		Pazopanib	Interferon-alpha	£32,062	0.717	£44,697
		Pazopanib	Best supportive care	£36,356	0.979	£37,126
SMC	2007	Sunitinib	Interferon-alpha	Not reported	Not reported	£33,371
Thompson Coon	2010	Sunitinib	Bevacizumab plus interferon-alpha	Not reported	Not reported	£171,301

		Sunitinib	Interferon-alpha	Not reported	Not reported	£71,462
NICE	2019	Nivolumab + ipilimumab	Sunitinib	Not reported	1.75	£28,068.31
			Pazopanib	Not reported	1.75	£28,021,92

Adapted from CS Appendix G, Table 3.

The ERG conducted ad hoc searches and these yielded a study of interest not captured in the company's searches.[43] This is likely to be because the study was published online in June 2019, after the company had conducted its last search. The study is a cost-effectiveness analysis of pembrolizumab plus axitinib versus sunitinib in first-line advanced RCC in China. It concludes that pembrolizumab plus axitinib was not likely to be cost-effective at a threshold of US$29,306 per QALYs gained. Although the study is not based on a UK perspective, it estimates mortality risk based on the OS curve of the KEYNOTE-426 trial also used to inform the company's model in the current appraisal.

In the CS, the company has narrowed down its focus to five previous NICE technology appraisals that are considered the most relevant comparators. These are TA169,[37] TA215,[44] TA 512,[27] TA542[24] and TA581[41] for sunitinib, pazopanib, tivozanib, cabozantinib and nivolumab plus ipilimumab respectively. The features of the models informing these TAs are summarised in comparison with the company's model in CS Table 41. Nivolumab plus ipilimumab is not listed as a comparator in the CS but nivolumab is featured as a second line treatment.

ERG conclusion

The ERG considers the company's search strategies and study selection criteria are robust and relevant to the decision problem.

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

Table 29 shows that the company's economic evaluation adheres to the NICE reference case requirements.

Table 29 NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per the scope developed by NICE	Yes	CS Table 1, page 10.
Comparator: As listed in the scope developed by NICE	Yes	Discussed in section 4.3.4.
Perspective on costs: NHS and PSS	Yes	Not explicitly stated in CS
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	Outcomes as per NICE scope (CS Table 1)
Type of economic evaluation: Cost utility analysis with fully incremental analysis	Yes	Cost utility analysis
Synthesis of evidence on outcomes: Based on a systematic review	Yes	Systematic literature review conducted to identify RCT relevant to submission. (CS section B.2.2).
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	Time horizon of 40 years.
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	EQ-5D data collected in company's trial.
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% pa for costs and health effects	Yes	

4.3.2 Model structure

The model structure, described in CS B.3.2 and illustrated in Figure 16, is reproduced in Figure 2 below. It is a partitioned survival 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 first-line treatments. At disease progression, patients

transition to the PD state, which is irreversible, so patients cannot return from PD to PFS. Patients in PF and PD states may die from cancer or other causes.

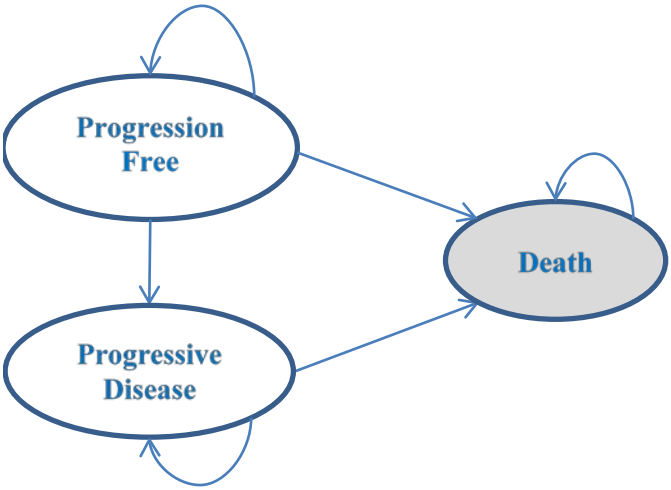

Figure 2 Structure of economic model

Reproduced from CS B.3.2 Figure 16

As mentioned above, all patients start in the PF state. At the end of each model cycle, patients may transition into a different state as estimated using a partitioned survival approach which is underpinned by the PFS and OS curves. The PFS curve estimates the proportion of patients who are progression free and is constrained by the OS curve i.e. the PFS cannot exceed the OS curve at any time point. The OS curve estimates the proportion of patients alive at each time point, while patients in the PD state are calculated as the remaining patients who are not dead and have progressed.

The submitted model includes analyses for two patient populations:

- The overall population of KEYNOTE-426;
- Subgroup population of patients with intermediate / poor RCC risk status in the KEYNOTE-426 population.

The PFS, OS and TTD curves for pembrolizumab plus axitinib, and sunitinib were based upon survival data from the KEYNOTE-426 trial. Sunitinib is assumed clinically equivalent to tivozanib and pazopanib. For the subgroup population of intermediate / poor RCC risk, pembrolizumab plus axitinib is compared to to sunitinib, pazopanib and tivozanib using survival data from the KEYNOTE-426 trial and is compared to cabozantinib using effect estimates from the company's

NMA, as no head-to-head comparison was available (NB. Cabozantinib is recommended only in patients at intermediate / poor RCC risk)

Subsequent treatment is incorporated in the model for some patients at the time of disease progression (described in more detail in section 4.3.7). Costs and utilities are applied to each of the health states (described in more detail in section 4.3.6 and 4.3.7).

The company's model also includes the following features and assumptions:

- **Cycle length**: 1 week with half cycle correction implemented.
- **Perspective**: NHS and PSS
- **Time horizon**: 40 years in the base case
- **Discounting**: 3.5% per year for costs and QALYs
- **Duration of treatment effects**: based on extrapolation of PFS and OS curves fitted to trial data and clinical expert judgement. A persistence of treatment effect throughout the model time horizon was assumed in the company's base case. Treatment waning after 10 years was tested in a scenario.
- **Adverse events**: includes grade 3 and above all-cause adverse events which occur in at least 5% of patients for all first-line treatments. Adverse events related to subsequent treatments are not explicitly modelled.
- **Utility and QALY calculations**: HRQoL estimates evaluated from the KEYNOTE-426 trial are used in the model. The company base case uses the estimation of utilities based on time-to-death. Two other approaches were used to estimate quality of life: estimation of utilities based on progression-free and progressed disease states (with or without differentiation by treatment). These approaches are discussed in detail in ERG Section 4.3.6. An age-based utility decrement is also applied.
- **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; monitoring and disease management in PF and PD states; treatment of included TEAEs for first-line treatments; and terminal care costs in the last cycle before death.
- **Uncertainty**: the model incorporates 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.

ERG conclusion

The three-state partitioned survival model is a standard modelling approach and has been applied in previous NICE appraisals for untreated advanced RCC. We consider that the model structure and partitioned survival approach is appropriate. The use of a 40 years model time horizon estimates lifetime costs and benefits, given the starting population age in the model. The company's model also includes an adjustment for age-related increase in mortality in the general population, by capping the projected OS curves to general population mortality rates.

4.3.3 Population

The model uses a cohort in the economic evaluation based upon the overall patient population of the KEYNOTE-426 trial. The key characteristics of patients included in the model are shown in Table 30 (CS Table 40). Expert clinical advice to the ERG is that patients starting first-line treatment for advanced RCC are slightly older than patients in the trial. (See section 3.1.3 of this report for more detail on the patient characteristics in the trial).

Table 30 Patient population characteristics in the model

Baseline characteristics	Model values
Age (years)	61.5
Male	71.3%
Patient weight (Kg)	81.5
Favourable RCC risk	Not explicit
Intermediate RCC risk	
Poor risk RCC	

The model also estimates cost-effectiveness for a subgroup of patients with intermediate or poor risk by IMDC classification. This subgroup was not specified in the NICE scope for this appraisal, but as explained earlier in this report, we consider this a clinically meaningful subgroup for assessment of cost-effectiveness.

4.3.4 Interventions and comparators

The economic model compares the cost effectiveness of pembrolizumab plus axitinib versus sunitinib, pazopanib, and tivozanib for the overall patient population, and compares against sunitinib, pazopanib, tivozanib and cabozantinib for the intermediate / poor RCC risk group. The ERG notes that the NICE technology appraisal of avelumab with axitinib for this indication is ongoing at the time of writing, and guidance is expected in early 2020. It is therefore not included as a comparator in the NICE scope or the company's decision problem.

In the base case analysis, tivozanib and pazopanib have been considered clinically equivalent to sunitinib. The company notes that this was accepted by the NICE appraisal committee in previous NICE appraisals for pazopanib, tivozanib, cabozantinib and nivolumab plus ipilimumab.

4.3.5 Treatment effectiveness and extrapolation

The company notes the follow-up period in KEYNOTE–426 was much shorter than the time horizon of the economic model and therefore extrapolation was necessary for OS, PFS and time on treatment (ToT) for the area-under-the-curve (AUC) partitioned survival approach. These extrapolations are discussed in more detail in this section for OS and PFS and ToT is discussed in section 4.3.7.

The company fitted parametric models to the KEYNOTE-426 KM data as recommended by the NICE DSU technical support document (number 14) on extrapolating survival data from clinical trials.[45] Firstly they estimated the goodness-of-fit statistics (i.e. Akaike information criterion [AIC] and Bayesian Information Criterion [BIC] and visual inspection of the agreement between the predicted and observed PFS, OS and ToT curves). Secondly, they examined the clinical plausibility of long-term extrapolations beyond the trial period.

4.3.5.1 Overall survival

The company assessed whether the proportional hazards assumption is reasonable by examining cumulative and log-cumulative hazard plots (CS Figure 19 and 20) for OS for pembrolizumab plus axitinib and for sunitinib from the KEYNOTE–426 trial. The company noted that the log–cumulative hazard plots of OS crossed and are not parallel and concluded that the proportional hazards assumption does not hold. Further, they concluded that the use of fully parametric modelling was most appropriate for extrapolation, as there were no abrupt changes in the log-cumulative hazard plots. The ERG agrees with the company's conclusions regarding

the proportional hazards assumption and notes that the methods used are consistent with the NICE DSU guidelines.[45] We provide our assessment of proportional hazards for PFS and OS in relation to the NMA earlier in this report (Table 8).

AIC and BIC statistics are shown for OS in CS Table 43. According to the AIC/BIC statistics, the best-fitting curve for pembrolizumab plus axitinib is the exponential, followed by the Gompertz. For sunitinib the best-fitting curve is the lognormal, following by the exponential. The company's clinical experts suggested that treatment with first-line sunitinib would be associated with 5 and 10 year survival between 20-25% and 10-15% respectively. The company's clinical experts suggested a five-year OS of approximately 50% when treated with pembrolizumab plus axitinib. One of the ERG's clinical experts suggested that the five-year survival of 50% for those treated with pembrolizumab plus axitinib may be optimistic.

The long-term OS predictions of pembrolizumab plus axitinib and sunitinib are shown in Table 31 and Table 32 (CS Appendix P Tables 3 and 4). On the basis of these predictions, the company concludes that the Gompertz, generalized gamma and lognormal distributions lead to clinical implausible outcomes.

Table 31 Long term OS predictions of pembrolizumab in combination with axitinib

Year	Exponential	Weibull	Log-logistic	Log-normal	Gompertz	Generalized Gamma
1	88.3%	88.6%	88.5%	88.3%	88.9%	88.7%
2	78.0%	76.2%	76.8%	79.2%	74.4%	75.6%
5	53.5%	44.9%	51.9%	62.4%	20.3%	38.5%
10	28.7%	16.5%	31.6%	47.6%	0.0%	6.2%
20	8.2%	1.7%	16.5%	31.0%	0.0%	0.0%

Reproduced from CS Appendix P Table 3

Table 32 Long term OS predictions for sunitinib

Year	Exponential	Weibull	Log-logistic	Log-normal	Gompertz	Generalized Gamma
1	79.9%	80.1%	79.7%	79.5%	79.7%	79.3%
2	63.9%	62.6%	63.6%	65.5%	65.5%	66.8%
5	32.5%	28.2%	37.3%	43.6%	42.0%	48.4%
10	10.6%	6.9%	20.9%	27.9%	27.6%	35.4%
20	1.1%	0.3%	10.5%	15.5%	17.9%	22.8%

Reproduced from CS Appendix P Table 4

The company compared the long-term OS predictions (exponential) of sunitinib against published study estimates, for external validation (CS Figure 23). The ERG compares the long-term OS predictions of sunitinib using the exponential, Weibull and Log-logistic with the trial with the longest follow-up, i.e. the COMPARZ trial (Figure 3).

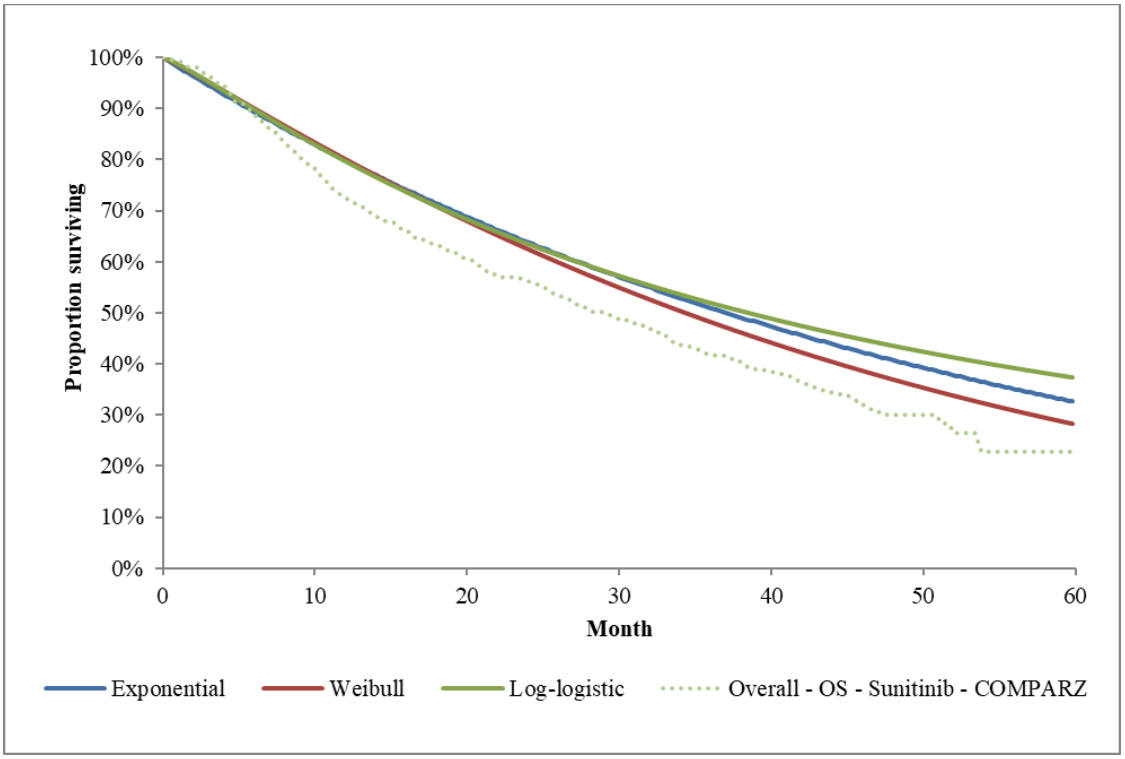

Figure 3 Modeled OS vs. selected OS external validation source for sunitinib

The company chose the exponential as the most appropriate distribution to extrapolate OS for the sunitinib arm. They justify this by stating that the log-cumulative hazard plots show a constant hazard over time suggesting the exponential is appropriate, the AIC and BIC showed close statistical fit to the observed data, the exponential distribution provides long-term OS estimates expected to be seen with sunitinib according to external data and in line with estimates from clinical experts.

The company chose the log-logistic as the most appropriate distribution to extrapolate OS for the pembrolizumab plus axitinib arm. They justify this choice on the basis the AIC/BIC showed a good statistical fit to the observed data and the tail of the log-logistic curve was considered by clinical experts to be more credible based on the expectation that a percentage of patients would derive a long-term survival benefit from the combination of an immunotherapy with a

tyrosine kinase inhibitor. This immunotherapeutic effect would imply a declining, rather than a constant hazard over the long term.

The company notes that NICE technical support document 14 states that "While fitting separate parametric models to individual treatment arms may be justified, it is important to note that fitting different types of parametric model (for example a Weibull for one treatment arm and a log normal for the other) to different treatment arms would require substantial justification".[45] The company's justification is that the mode of action of combination of immunotherapy with a TKI is not comparable to the mode of action associated with TKI monotherapy. The company also states that none of the parametric distributions gave clinically plausible long-term OS estimates for both arms simultaneously.

The ERG disagrees with the company's justification to use a different distribution for treatment arms due to a different mode of action of combination immunotherapy plus TKI to TKI monotherapy. The ERG notes that the OS survival data is immature and for pembrolizumab plus axitinib the data does not demonstrate an underlying hazard that is similar to the log-logistic. Furthermore, the underlying hazard is similar to sunitinib. Finally, the ERG notes that the NICE appraisal committee did not consider that the modelling of the immunotherapeutic effect was substantiated by evidence in TA581[41] for nivolumab plus ipilimumab and that it could not generalise the size of this effect from one cancer to another. It concluded "that there was no robust evidence on the size of the association between a clinically meaningful definition of response and long-term survival for nivolumab and ipilimumab"' The ERG considers the committee's decision is relevant to this current appraisal.

The ERG considers that both the exponential and Weibull distributions are plausible for OS for pembrolizumab plus axitinib and sunitinib and that the Weibull distribution provides a better fit for the long-term OS of sunitinib (Figure 4). Therefore, we have chosen this as the most appropriate distribution, although we caution that due to the immature OS data this choice may be somewhat speculative. The ERG base case analyses use the Weibull distribution for OS and are shown in section 4.4. We consider scenarios with exponential and log-logistic extrapolations for OS in section 4.4. We have also run scenarios using time varying hazard ratios in section 4.4.

The company has not included the assumption of treatment effect waning (reducing) in their base case. They justify this by the fact that waning of effect has not been included in previous NICE appraisals for RCC, and that patients continue to be treated with axitinib after the 2-year stopping rule for pembrolizumab. In addition, the company states that they believe a proportion of patients would derive a long-term survival benefit from the combination of an immunotherapy with a TKI. The ERG notes that a proportion of patients would receive second-line treatment after disease progression and this second-line treatment would influence their survival. Further, many patients who receive sunitinib as first line treatment would receive nivolumab as second-line therapy and so it may be the case that OS for patients receiving second-line treatment may be similar between treatment arms. However, as the OS data from KEYNOTE-426 is immature, we have only included treatment waning in a scenario analyses.

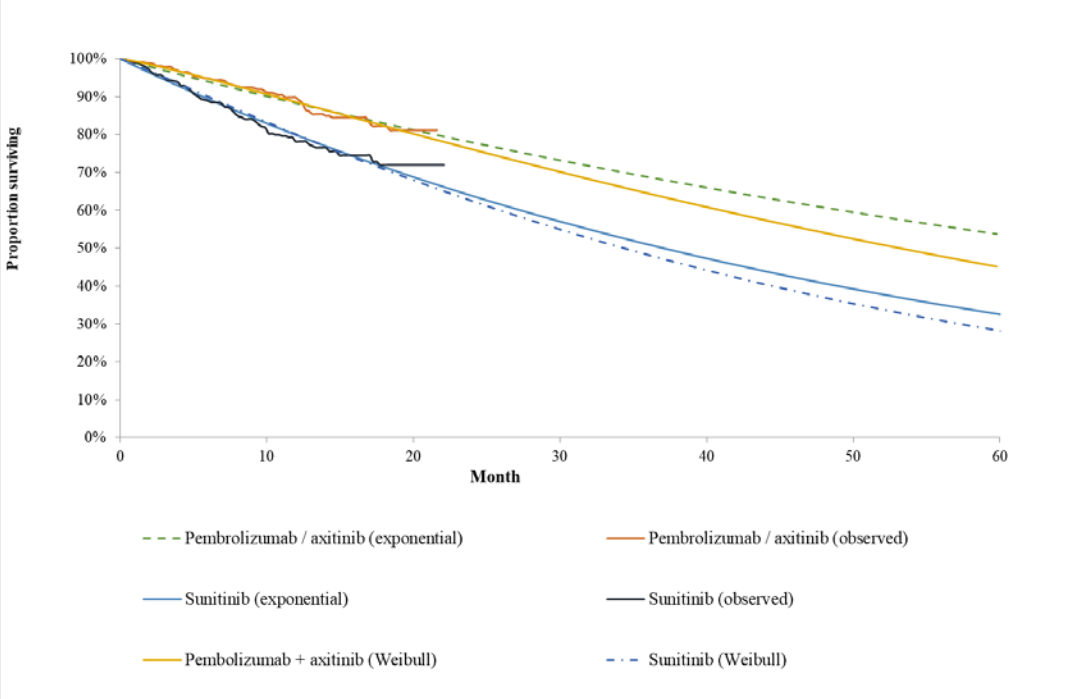

Figure 4 OS from KEYNOTE–426 compared to fitted curves for the exponential and Weibull distributions

4.3.5.2 Progression-free survival

The best AIC and BIC statistical fit for PFS for the pembrolizumab plus axitinib was the log-normal distribution, followed by the generalized gamma distribution. The best AIC/BIC statistical fit for the sunitinib arm was the exponential distribution, followed by the Weibull distribution. Based upon the AIC/BIC and visual fit, the company chose the exponential distribution for both treatment arms. The modelled PFS is compared against the observed data in CS Figure 31 (Figure 5).

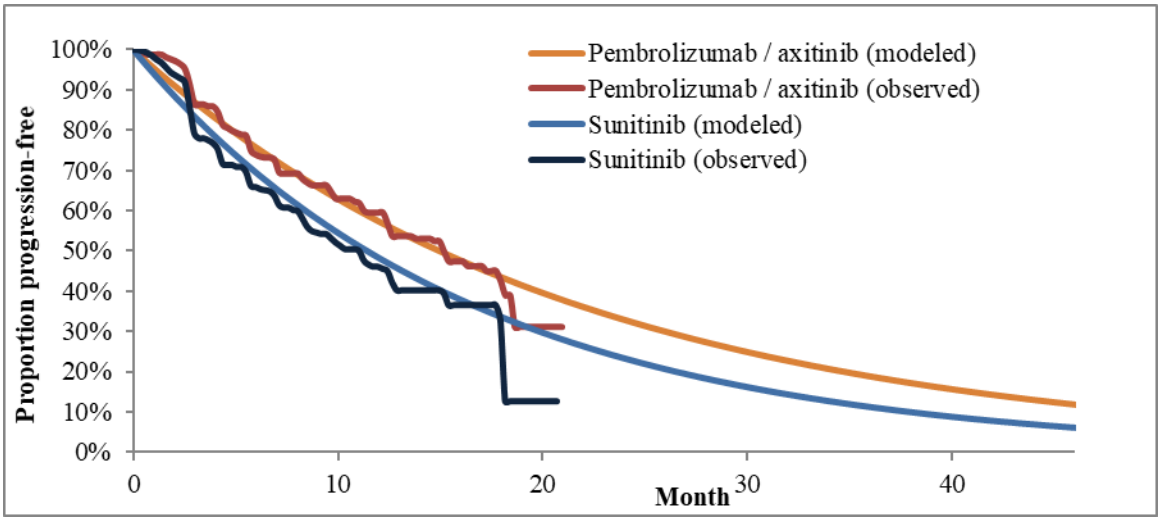

Figure 5 PFS KM curves vs fitted 2-phase piecewise model with cut off at 13 weeks and exponential distribution thereafter

The company noted that there was a steep increase in patients' disease progressing at 13 weeks due to the imaging being performed at 12 weeks. They have therefore used piecewise modelling whereby the KM data is used for the first 13 weeks and an exponential distribution is used thereafter. The ERG considers this approach to be appropriate as the exponential provides a good fit to the COMPARZ trial data[12] for PFS and a good statistical and visual fit. However, we note that if the KM data is to be used, it could be used for a longer time-period than for 13 weeks, such as for 54 weeks. We requested a scenario analysis from the company using the KM data for a longer time-period of 54 weeks (clarification question B15). The results were only marginally different. We conduct scenario analyses with Weibull and log-logistic extrapolations for PFS in section 4.4.

4.3.5.3 Intermediate / poor RCC risk subgroup analysis

The company compares pembrolizumab plus axitinib versus sunitinib pazopanib, tivozanib and cabozantinib for patients with intermediate or poor risk status. For this subgroup, the company fits curves for OS, PFS and ToT for pembrolizumab plus axitinib and sunitinib. The same distributions were used as described above for the base case analysis. The company provides more detail on the fitting process in response to a clarification question (B9). As for the overall RCC population, the ERG agrees with the company's choice for PFS for the intermediate or poor risk score population, but disagrees with the company's choice of the log-logistic for OS for pembrolizumab and axitinib. The ERG prefers to use the Weibull for both treatment arms for OS.

For cabozantinib, the model uses a time-constant HR for OS and PFS. The hazard ratio is taken from the company's NMA (PFS HR = [redacted] and OS HR = [redacted] vs. pembrolizumab plus axitinib (see section 3.1.7 of this report). For tivozanib and pazopanib, PFS and OS was assumed equivalent to that estimated by the sunitinib arm. For this to be clinically meaningful the survival estimates from the sunitinib arm of KEYNOTE-426 should be representative of estimates from pivotal phase III trials of sunitinib. In the phase III registration trial for sunitinib[23] the median PFS was 11 months, which is similar to the 11.1 months estimate from KEYNOTE-426. Expert clinical advice to the ERG is that the sunitinib PFS estimates from KEYNOTE-426 accord with what those seen in the earlier pivotal trial.

ERG conclusion

The methods used to extrapolate OS and PFS for the economic model are reasonable and consistent with NICE recommended methodology, although the ERG disagrees with the choice of curves chosen for OS. Parametric survival curves were fitted to both the pembrolizumab plus axitinib and sunitinib treatment arms for the KEYNOTE-426 trial. The trial data for OS are immature which makes the choice of a parametric curve extrapolating beyond the trial duration more uncertain. Further, the model results were very sensitive to changes in the parametric curve used for OS extrapolation. The company uses a log-logistic distribution for pembrolizumab plus axitinib and an exponential distribution of OS extrapolation. The ERG prefers the Weibull distribution for the OS extrapolation for pembrolizumab plus axitinib and for sunitinib as the Weibull distribution provides a better fit for the long-term OS of sunitinib, and in principle the same distribution should be used for both treatment arms. The company used the exponential distribution for the PFS extrapolation, and we agree with this choice.

4.3.5.4 Adverse events

Adverse events are included in the economic model for Grade 3+ all-cause AEs which occurred in at least 5% of patients (for any grade AE). Adverse event data for pembrolizumab plus axitinib and sunitinib are from the KEYNOTE-426 trial (CS Table 46 for Grade 3+). In response to clarification question B4, the company noted that some of the values in this table were incorrect. The correct values are shown in Table B4.1 of the clarification response document. The safety profile for tivozanib and pazopanib is assumed to be equal to the safety profile of sunitinib. The incidence AEs for cabozantinib was taken from the published data from that NICE TA542 (cabozantinib).[24]

In the base case, the impact of AEs was incorporated by estimating weighted average disutilities and costs per patient, as described in section 4.3.6 and 4.3.7.

4.3.6 Health related quality of life

The company conducted a systematic literature review to identify HRQoL (in terms of utilities) associated with metastatic RCC. The company search strategy is described in CS Appendix H. The company conducted an initial search on March 14, 2018, with an update search completed on 20th February 2019. The company searched EMBASE, MEDLINE, the Cochrane Central Register of Controlled trials and the Cochrane Database of Systematic Reviews. In addition, a targeted search was conducted in various relevant Oncological and Pharmacoeconomic conference proceedings. Further the company searched the grey literature including reports from NICE, Scottish Medical Consortium (SMC) and the Institute for Quality and Efficiency in Healthcare (IQWIG).

The eligibility criteria for the HRQoL studies included generic, disease-specific and preference-based outcome measures (Table 1 in CS Appendix H). The original search identified 25 full-text articles after abstract and full-text screening. The update search found a further seven studies. Of the studies identified, six studies reported utility values (EQ-5D) and these are shown in Table 3 in CS Appendix H. The ERG notes that none of these studies have been used in the economic model for scenario analyses.

EQ-5D-3L data were collected in the KEYNOTE–426 trial (see section 3.3.3 of this report) and these data were used in the economic model. The company states that the estimated utilities

used in the model were derived directly from patients and evaluated using UK-preference scores and this is consistent with the NICE reference case.[46] The ERG agrees that the utility values meet the NICE reference case and are suitable for inclusion in the model.

In KEYNOTE-426, for pembrolizumab plus axitinib, the EQ-5D questionnaire was administered on day one of every cycle from cycle one to nine; on day one of every other cycle from cycle 9 - 19; and on day one of every 4th cycle from cycle 19 until treatment discontinuation and the 30-day post-treatment discontinuation follow-up visit. Each cycle length was equal to 21 days. For sunitinib, the EQ-5D questionnaire was administered on days one and 29 of every cycle from cycle one to cycle four; on day one of every cycle from cycle 5 - 10; and on day one of every other cycle from cycle 10 until treatment discontinuation, and the 30-day post-treatment discontinuation follow-up visit. Each cycle length was equal to 42 days.[47]

A regression analysis consisted of EQ-5D data from 850 individuals. CS Table 47 shows the level of compliance at difference time points, i.e. those who completed the EQ-5D questionnaire. The company analyses the data according to treatment, disease progression, time to death and adverse events. For the company's base case analysis, they used time-to-death utility data, where utility data is estimated for the time-period until death (Table 33). They stated that this approach had been used in NICE appraisals for patients with advanced non-small cell lung cancer who had previously received platinum-based chemotherapy or palliative radiotherapy[48] and in advanced melanoma patients.[49] The utility values based on time-to-death are shown in Table 33 (CS Table 50).

Table 33 EQ-5D health utility scores by time-to-death

	Pooled (N=532), number of observations: 2,704		
	Estimate	**SE**	**95% confidence interval**
≥360 days	[redacted]	[redacted]	[redacted]
180 to 360 days	[redacted]	[redacted]	[redacted]
90 to 180 days	[redacted]	[redacted]	[redacted]
30 to 90 days	[redacted]	[redacted]	[redacted]
0 to 30 days	[redacted]	[redacted]	[redacted]
AE disutility	[redacted]		

Reproduced from CS Table 50

Whilst it has been more common in previous technology appraisals to use health state specific utility values, rather than time to death, Hatswell et al[50] noted that disease progression may not fully capture all predictive factors of patient utility and time-to-death provide a good fit to patient data. The company conducted scenario analyses using treatment-specific health-state based utilities and the pooled health state-based utilities from KEYNOTE–426 (CS Table 67). These scenarios only produced a small change in the cost-effectiveness results. The ERG therefore considers either approach to be reasonable.

In response to ERG clarification question B12, the company provided treatment-specific time to death-based utilities (shown in Table B12.1 of the clarification response document). The ERG notes that the utility values for patients on each treatment are not statistically different from each other and so agrees that is appropriate to assume the same utility values for patients who start on pembrolizumab plus axitinib or sunitinib. The ERG notes that the utility values for patients with ≥360 days until death is higher than the UK population norm for this group. According to Kind et al,[51] the weighted average utility of men and women at this age group is 0.775. We conducted a scenario analysis where the utility value for patients with ≥360 days until death is set to 0.775 (see section 4.4).

The ERG agrees with company's approach to evaluating health utilities. We conducted scenario analyses using the utility values from previous NICE TAs for tivozanib and pazopanib (section 4.4).

4.3.6.1 Adverse event disutilities

Adverse event disutilities were estimated according to the EQ-5D values collected in the KEYNOTE-426 trial for pembrolizumab plus axitinib and sunitinib. These estimates differed according to the method used for calculating the utility values: for the progression status model, this disutility was calculated as [redacted], for the treatment-specific progression status model this disutility was calculated as [redacted] and for the time-to-death utility model, this was calculated as [redacted].

The mean duration of the AE was estimated from KEYNOTE–426, according to the specific AE. This mean duration was applied together with the disutility associated with AEs and the overall incidence rates of AEs to estimate a one-off QALY loss per patient for each treatment ([redacted] for

pembrolizumab plus axitinib and [redacted] for sunitinib). The QALY losses were applied to the first cycle of the model for each treatment arm only.

4.3.6.2 Age-related disutility

The company includes age-related disutility using the formula provided by Ara and Brazier, [52] reweighted using the starting age in the model of 62.5 years. The ERG notes that including age-adjusted utility is recommended by NICE DSU Technical Support Document 12.[53] In response to clarification question B11, the company analyses the effect of age on utility values. The company found that utility values were not associated with age. The ERG suggests that age-related disutility should not be included in the model. The ERG base case analysis therefore does not include an age-related disutility (section 4.4).

ERG conclusion

The company's approach to estimating utility values is reasonable and consistent with the NICE reference case. The use of KEYNOTE-426 utility data is preferable to other sources.

4.3.7 Resource use and costs

The costs included in the economic model consist of drug acquisition and administration for first and subsequent treatments, health state management cost, costs for managing AEs and terminal care costs incurred at the end of life.

The company conducted a comprehensive literature search to identify costs and resource used in the treatment and management of advanced renal cell carcinoma patients. The original search was completed on 14th March 2018, with an update search on 20th February 2019. The search was limited to those studies published after 1st January 2007. Details of the search strategy and eligibility criteria are shown in CS Appendix G. Studies were only included if they reported UK costs and resource use of metastatic renal cell carcinoma from a UK perspective.

After abstract and full-text screening, nine studies were identified and this was increased to 10 studies when the NICE appraisal of nivolumab and ipilimumab for untreated advanced renal cell carcinoma was published. The ten included studies are shown in CS Appendix I. The ERG considers that the company's literature review is likely to reflect the available evidence.

4.3.7.1 First-line drug acquisition costs

The cost per pack for all drugs are taken from the British National Formulary.[54] Dosages are taken from each treatment's Summary of Product Characteristics. Intended dosages were adjusted by the dose intensity observed in the treatments' trials. None of the treatments for first-line or subsequent treatment lines are eligible for vial sharing.

Pembrolizumab is administered as a 30-minute IV infusion of 200mg every three weeks. The list price of a 100mg vial is £2,630. Patients treated with pembrolizumab are treated until disease progression or unacceptable toxicity. There is a stopping rule for pembrolizumab such that patients do not receive treatment with pembrolizumab beyond 24 months. Axitinib is administered twice daily as an oral treatment with a fixed dose of 5mg. The list price of a packet of 56 tablets of axitinib is £3,517. A course of treatment has a four week cycle length. Patients may continue treatment with axitinib beyond 24 months. Pembrolizumab and axitinib are supplied to the NHS with a commercial access agreement and a confidential patient access scheme (PAS) respectively.

The dosing, frequency and unit costs of the first-line drugs are shown in Table 34 (CS Table 52). Several treatments are available with confidential patient access schemes (PAS). The company has reported all analyses using the list price of the treatments. The ERG has replicated the company's analyses using the treatment PAS prices in a separate confidential appendix to this report.

Table 34 Dosing, frequency and unit costs per administration for intervention and comparator

Drug	Dosing Schedule	Frequency of admin-istration	Total dose required per admin (mg)	Cost per administration (assuming no wastage)	Dose intensity	Cost per admin-istration (list price)
Pembrolizumab	200 mg IV Q3W	Q3W	200	£5260.00	[redacted]	£4,986.48
Axitinib	5 mg BID orally	Q4W	280	£3,517.00	[redacted]	£2,975.38
Sunitinib	50 mg QD orally for 4 weeks, then 2 weeks off treatment	Q6W	1,400	£3,138.80	[redacted]	£2,344.68
Pazopanib	800 mg QD orally	Q4W	22,400	£2,092.53	86.0%	£1,799.58

Tivozanib	1.34 mg QD orally for 3 weeks followed by 1 week without treatment	Q4W	28	£2,052.00	94.0%	£1,928.88
Cabozantinib	20/40/60 mg QD orally	Q4W	1,680	£4,800.13	94.3%	£4,526.53

Reproduced from CS Table 52

4.3.7.2 Time on treatment

Parametric curves were fitted to the patient level treatment duration data from KEYNOTE-426. AIC/BIC statistical tests indicated that the best fit was the Weibull for pembrolizumab, log-normal for axitinib and log-normal for sunitinib (CS Table 54). However, the company's clinical expert estimated that about 5-10% of patients would still be receiving sunitinib after 5 years, whilst the log-normal estimated 12% would be receiving treatment. Hence the log-normal was considered implausible. The company chose the exponential distribution for consistency with PFS for axitinib and sunitinib. For pembrolizumab, the Weibull curve was chosen as it had the best statistical and visual fit.

For the subgroup analysis, for the intermediate/poor risk group, the log-logistic distribution was used for pembrolizumab based on visual inspection and AIC/BIC. Exponential ToT curves were used for axitinib and sunitinib. For cabozantinib, the proportion of patients remaining on treatment was based on the modelled PFS curve for this treatment arm.

For first-line treatment, the company's base case does not cap the ToT curves with the PFS curves, meaning that patients could potentially continue to receive treatment even after they have progressed. The ERG observed that PFS and ToT curves are similar and therefore the company's choice to not cap ToT is not likely to drive model results.

There is no waiting period between stopping first-line treatment and starting second-line treatment in the company's model. Patients who progress are assumed to immediately commence second-line treatment. This was considered reasonable by the ERG's clinical experts.

ERG conclusion

The ERG agrees that the log-normal would be an implausible distribution for ToT for axitinib and sunitinib. As described in section 4.3.5 for OS, the ERG considered that the company should use the same distribution for both treatment arms. This is also the case for ToT. Therefore, the ERG considers that the Weibull distribution provides the best visual fit for ToT for sunitinib and pembrolizumab, and it is also a good fit to the company's clinical expert estimate of patients remaining on treatment with sunitinib after 5 years (5 - 10%).

For the intermediate / poor risk subgroup analysis, the ERG prefers to use the same distributions for pembrolizumab plus axitinib and sunitinib. The Weibull appears to provide the best visual fit to the observed data.

4.3.7.3 Second-line treatment use and costs

The company includes second-line treatment costs according to two methods: real-world based and trial based. In the base case the company assumes that upon disease progression patients incur the costs of subsequent therapies in line with the NHS England submission in TA581 for nivolumab and ipilimumab in untreated RCC.[41] In this option, 50% of patients who had progressed were assumed to receive second-line treatment. The distribution of subsequent therapies is shown in Table 35 (CS Table 58).

Table 35 Type and distribution of second line subsequent treatments used in the base case

		First-line treatment				
		Pembrolizumab + axitinib	**Sunitinib**	**Pazopanib**	**Tivozanib**	**Cabozantinib**
Distribution of subsequent therapy	No active treatment	50.00%	50.00%	50.00%	50.00%	50.00%
	Pazopanib	30.00%	0.00%	0.00%	0.00%	0.00%
	Sunitinib	20.00%	0.00%	0.00%	0.00%	0.00%
	Nivolumab	0.00%	30.00%	30.00%	30.00%	30.00%
	Cabozantinib	0.00%	12.50%	12.50%	12.50%	0.00%
	Axitinib	0.00%	7.50%	7.50%	7.50%	7.50%
	Lenvatinib/ everolimus	0.00%	0.00%	0.00%	0.00%	12.50%*

	Source	NHS England Submission in TA581 *Assumption that the proportion of patients treated with cabozantinib in first-line that are expected to receive second-line treatment with cabozantinib were redistributed to lenvatinib/everolimus

Reproduced from CS Table 58

In the model, the proportion of patients estimated to progress in each treatment cycle is distributed between the six active second-line treatments and no treatment. The mean treatment durations from the trials of each treatment are then applied to each of the second-line treatments.[41, 55-65]

In a scenario analysis, the company assumes the same proportion of patients receiving subsequent therapy after disease progression as observed after progression in the KEYNOTE-426 trial. However, the company notes that some of the treatments are not recommended in UK clinical practice. The distribution of subsequent therapies in KEYNOTE–426 are shown in Table 36 (CS Table 59).

Table 36 Type and distribution of second line subsequent chemotherapies used in the base case

		First-line treatment				
		Pembrolizumab + axitinib	**Sunitinib**	**Pazopanib**	**Tivozanib**	**Cabozantinib**
Distribution of subsequent therapy	No active treatment	[redacted]	[redacted]	74.74%	74.74%	39.24%
	Axitinib	[redacted]	[redacted]	11.45%	11.45%	23.08%
	Cabozantinib	[redacted]	[redacted]	0.00%	0.00%	1.28%
	Everolimus	[redacted]	[redacted]	10.18%	10.18%	8.98%
	Lenvatinib / everolimus	[redacted]	[redacted]	0.00%	0.00%	1.28%
	Nivolumab	[redacted]	[redacted]	0.00%	0.00%	10.77%
	Pembrolizumab	[redacted]	[redacted]	0.00%	0.00%	7.18%
	Sunitinib	[redacted]	[redacted]	0.00%	0.00%	14.11%
	Temsirolimus	[redacted]	[redacted]	0.00%	0.00%	8.98%
	Pazopanib	[redacted]	[redacted]	0.00%	0.00%	17.95%
	Cytokines (interferon)	[redacted]	[redacted]	8.90%	8.90%	3.85%
	Source	KEYNOTE-426 trial	KEYNOTE-426 trial	Assume equal to Tivozanib	TIVO-1 trial	CABOSUN trial

Reproduced from CS Table 59

The costs of each subsequent treatment are detailed in Table 37 (CS Table 60). In all cases drug costs have been sourced from the BNF,[54] and applied to dosing regimens as per each therapy's SmPC. Median treatment duration was taken from the relevant trials for each of the treatment and then converted to mean treatment duration by assuming constant hazards.

Table 37 Subsequent therapy- drug formulation, dose, administration, mean treatment duration and total drug acquisition cost

Subsequent treatment	Dosing schedule	Drug acquisition cost per admin (2018 GBP)	Mean treatment duration (months)	Total drug acquisition cost (2018 GBP)[a]
Nivolumab	480 mg IV Q4W or 240 mg IV Q2W	4,846.56	7.9*	41,804.38
Pembrolizumab	200 mg IV Q3W	4,986.48	7.9	57,348.36
Axitinib	5 mg orally BID	3,587.34	11.8*	46,133.02
Cabozantinib	60 mg orally QD	4,800.13	12.1*	63,235.08
Lenvatinib / everolimus	18 mg orally QD	1,810.62	11.0*	42,856.12
	5 mg orally QD	1,785.00	11.0*	n/a
Pazopanib	800 mg orally QD	1,574.63	10.7*	20,884.69
Sunitinib	50 mg orally QD for 4 weeks, then 2 weeks off treatment	2,344.68	10.7*	18,140.54
Everolimus	10 mg orally QD	2,290.23	6.3*	15,803.62
Temsirolimus	25 mg IV QW	103.58	6.3*	13,177.12
Cytokines (Interferon a2B Roferon-A)	10 MU SC three days per week	1,345.20	4.0*	2,458.35

** Mean ToT was calculated as a function of median ToT, based on an assumption of constant hazards.*
Key: BID, twice daily; IV, intravenous; BNF, British National Formulary; Q2W, once every 2 weeks; MG, milligrams; MU, million units; SC, subcutaneously
[a] Values corrected in company's clarification response (B1). Values correct in model so no changes to company's base case results.

Adapted from CS Table 60

The ERG received advice from its clinical experts that the proportion of patients receiving subsequent therapy was between 55% - 70% of those who had progressed after first line treatment. However, NHS England were clear in their submission for NICE TA581[41] that a second line treatment rate of 50% is appropriate in 2018. Further, clinical experts to the ERG advised that more patients would receive cabozantinib after first line treatment with pembrolizumab plus axitinib than the estimates shown in Table 35. The ERG has therefore made changes to the proportion of patients receiving cabozantinib as a second-line treatment for the ERG base case. An alternative scenario is also run where a high proportion of patients receive second-line treatment (section 4.4).

4.3.7.4 Treatment administration costs

The company includes treatment administration costs. Pembrolizumab is administered as a IV infusion (30 minute administration). The administration cost of £174.40 is taken from the National Schedule of Reference Costs (currency code SB12Z).[68] The other first-line treatments are oral chemotherapies and incur an administration cost of (SB11Z). The administration costs of first-line treatments are shown in CS Table 55.

We note that there does not need to be an administration cost for oral chemotherapies in the model. As reported in TA542,[24] the ERG noted that the NHS does incur costs for the delivery of oral chemotherapies. However, the modelled health state costs include a monthly consultant-led medical oncology outpatient visit and blood tests, which was assumed to include the cost of procurement, prescribing and monitoring of oral chemotherapies. The ERG has reduced the administration cost of oral treatment to zero in the ERG base case (section 4.4).

Administration costs for second-line treatment are shown in CS Table 61. The ERG notes that the administration costs for nivolumab were incorrectly reported as £309.20, whereas the value used in the model is £174.40, i.e. the same value as for pembrolizumab.

4.3.7.5 Health state unit costs

The resource use and unit costs of progression-free and progressed disease are shown in Table 38 (CS Table 56). The unit costs were taken from the latest NHS reference costs.[68] The company states that the resources for the health states were based upon those from the NICE TA542.[24]

The health state costs are £51.05 per weekly cycle for the progression-free and progressed health states. An additional cost of £229 was applied in the first cycle of the model for the first attendance outpatient consultation.

The ERG considers that the company's estimates of health state costs are reasonable. They reflect resource use assumptions in previous NICE appraisals for untreated RCC (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.[68]

Table 38 Resource use and unit costs of progression-free, progressed and terminal health states within the model

	Resource	Resource use (per cycle)	Reference	Unit cost	Reference[68]
PFS	Outpatient consultation (first attendance)	N/A	NICE TA542	£229.00	NHS reference costs 2017-2018 Currency code WF01B, Service code 370, Medical oncology
	Outpatient consultation (follow-up attendance)	0.25		£166.00	NHS reference costs 2017-2018 Currency code WF01A, Service code 370, Medical oncology
	CT Scan	0.08		£110.00	NHS reference costs 2017-2018 Currency code RD25Z Computerised Tomography Scan of three areas, without contrast
	Blood test	0.25		£3.00	NHS reference costs 2017-2018 Currency Code: DAPS05
	Total cost per week	**Cycle 1: £280.05**		**Subsequent Cycles: £51.05**	
PPS	Outpatient consultation (follow-up attendance)	0.25	NICE TA542	£166.00	NHS reference costs 2017-2018 Currency code WF01A, Service code 370, Medical oncology
	CT Scan	0.08		£110.00	NHS reference costs 2017-2018 Currency code RD25Z Computerised Tomography Scan of three areas, without contrast
	Blood test	0.25		£3.00	NHS reference costs 2017-2018 Currency Code: DAPS05
	Total cost per week		**Every cycle: £51.05**		

Reproduced from CS Table 56

4.3.7.6 Cost of terminal care

The company includes a cost of terminal care of £6,789.76 based upon a previous HTA submission for this disease[24] and inflating to 2017/8 prices. The CS notes that there is limited data for the cost and resource use of RCC patients in terminal care. The cost of terminal care is assumed to be the same for all treatment arms.

The ERG for the cabozantinib TA542[24] considered the cost used was an underestimate of the actual costs of terminal care, 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, they 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.[69] However this is for the year 2016-17. We have used a similar methodology to update that cost to £8073 for 2017/8. We include this revised figure in ERG analyses in section 4.4.

4.3.7.7 Adverse event costs

The model includes the costs of managing grade 3+ adverse events. The resources used for the management of adverse events was mainly derived from previous technology appraisals for untreated advanced or metastatic RCC[24, 41] or metastatic urothelial carcinoma.[70] Unit costs were taken from the latest NHS reference costs 2017/8.[68] The unit costs of the management of adverse events and the assumptions used are shown in Table 39 (CS Table 57).

The unit cost of treating diarrhoea is higher than the value used in previous TAs, for example the unit cost of treating diarrhoea in TA581[41] is £788.25. However, changes to the unit cost of treating diarrhoea has minimal effect on the cost effectiveness results.

4.3.8 Model validation

The company states that their modelling approach was validated externally by the University of Sheffield's School of Health and Related Research (SCHARR) with input by two external health economists (CS section B.3.10). It further states that details of this validation include model structure, selection of appropriate datasets, survival analysis undertaken and assumptions surrounding extrapolation of survival, quality of life and healthcare resource use. The CS further states that quality assurance internal validation was carried out by the economists who produced the economic model and no major errors were found.

Below is a list of verification checks undertaken by the ERG. These include checks on input data and technical validation of coding.

Table 39 Unit costs of adverse events

Grade 3+ AE with incidence >5%	Unit Cost	Reference	Rationale
Alanine aminotransferase increased	£0.00	Based on the assumption: Regular blood tests (already considered under health-state management costs)	TA542[24]
Aspartate aminotransferase increased	£0.00	Based on the assumption: Regular blood tests (already considered under health-state management costs)	TA542[24]
Decreased appetite	£615.76	Non-elective short stay	TA581[41]
Diarrhoea	£1248.34	Non-elective short stay	TA581[41]
Fatigue	£657.76	Non-elective short stay, cost of face to face community nurse	TA581[41]
Hyperglycaemia	£156.00	Based on the assumption of 1 visit to endocrinologists, initiation of therapy with anti-diabetic medication: metformin 500mg one daily for one year	TA542[24]
Hypertension	£850.21	Non-elective short stay, consultant medical oncology visit WF01A; *non-admitted face to face attendance, follow-up*, 2 follow up GP visits	TA519[70]
Hyponatremia	£0.00	Based on the assumption: Regular blood tests (already considered under health-state management costs)	TA542[24]
Lipase level increased	£357.13	Regular day and night admission SA04J *Iron deficiency Aneamia with CC score 6-9*	TA581[41]
Lymphocytopenia	£331.90	Assumed that 20% of short stay emergency tariff (weighted average of SA25A-SA35E) and 80% of patients with day case tariff (weighted average of SA35B-SA35E)	TA542[24]
Neutropenia	£80.50	Assumed that 10% of patients require hospital treatment, each requiring two episodes during therapy. Weighted average of mean costs for HRG code WJ11Z *Other disorders of immunity* across non-elective long- and short-stay episodes and day-case admissions	TA519[70]
Neutrophil count decreased	£80.50	Assumed to be equal to neutropenia	TA519[70]
Palmar-plantar erythrodysaesthesia syndrome	£615.76	Non-elective short stay	TA581[41]
Platelet count decreased	£80.50	Assumed to be equal to neutropenia	TA519[70]
Stomatitis	£615.76	Non-elective short stay	TA581[41]
Thrombocytopenia	£357.13	Regular day and night admission SA04J *Iron deficiency Anaemia with CC score 6-9*	TA581[41]

Reproduced from CS Table 57

4.3.8.1 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 manually ran scenarios checking all model outputs (for both the IMDC risk subgroup and the overall risk population) against results reported in the CS for the base case, PSA and DSA results.
- We traced input parameters from entry cells in the model ('Raw' inputs sheets), to PSA / DSA sampling (on the "DSA Results" and "PSA Setup" sheets) 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 and wastage), health state costs and adverse event costs and QALY loss;
- Survival curve calculations were checked ("Effectiveness_survival" sheet, all the treatment effectiveness sheets and "ToT_Parametric Estimation" sheet.
- We estimated cohort sizes in the three states at each cycle using alternate but corresponding formulas.
- We checked QALY and cost calculations on the Markov sheets for all treatments.

ERG conclusion

We spotted a few inconsistencies in parameter values between the CS and the company's model. In response to ERG clarification questions the company states that the values in the model are correct and therefore do not affect the results reported in the CS or the model outputs. The ERG did not spot any errors in the Excel spreadsheet formulas of the company model.

4.3.8.1 Assessment of internal and external validity of model

The company's fitted survival curves are described in detail earlier in this report (section 4.3.5). In the base case, these curves are based on the results of the KEYNOTE-426 trial. In general, the parametric curves chosen by the company provide a good fit to the observed data for PFS, OS and ToT.

The ERG assesses the external validity of the model by comparing mean life years for patients treated with sunitinib with those from previous NICE technology appraisals. The results are shown in Table 40. The mean life years for sunitinib vary between 2.845 – 4.53 years depending on the assumptions used to extrapolate OS. The ERG estimate of mean life years for sunitinib is similar to the ERG estimates in previous NICE appraisals.

Table 40 Mean life years for sunitinib in the current appraisal compared to previous TAs

	Mean life years for sunitinib		
	Current appraisal	TA581[41]	TA512[27]
Company's estimate	3.86	4.53	2.846
ERG's estimate	3.47	3.03	3.31

4.3.9 Cost effectiveness results

Results from the economic model are presented in section B.3.7, page 136 of the CS as incremental costs per QALY gained for pembrolizumab plus axitinib compared against sunitinib, tivozanib and pazopanib. Results are also presented in terms of life years gained. For the overall population, the results are presented pairwise against pembrolizumab plus axitinib for all comparators, with pazopanib and tivozanib assumed to be clinically equivalent in effect to sunitinib. Two sets of pairwise base case results are presented in the CS: Table 64 which presents a comparison with sunitinib and CS Table 65 which presents a pairwise comparisons with tivozanib and pazopanib. These tables are reproduced below in Table 41.

Table 41 Base case cost effectiveness results for the overall patient population

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) versus baseline (QALYs)
Pembrolizumab + axitinib	[redacted]	6.887	[redacted]	-	-	-
Sunitinib	[redacted]	3.864	[redacted]	£137,537	2.320	£ 59,292
Pazopanib	[redacted]	3.864	[redacted]	£133,472	2.320	£ 57,540
Tivozanib	[redacted]	3.864	[redacted]	£131,402	2.320	£ 56,648
ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years						

Adapted from CS Table 64 and CS Table 65

For the company base case, there is an incremental cost effectiveness ratio (ICER) £59,292 per QALY for pembrolizumab with axitinib compared to sunitinib. ICERs of £57,540 and £56,648 are reported for pembrolizumab plus axitinib compared to pazopanib and tivozanib respectively.

Base case results are also reported for the intermediate/poor risk subgroup. Pairwise cost-effectiveness results are presented for pembrolizumab plus axitinib compared to sunitinib, pazopanib, tivozanib and cabozantinib with the assumption that pazopanib and tivozanib both have equivalent clinical efficacy to sunitinib (CS Tables 68 and CS Table 70). The pairwise results are £59,766, £58,350, £57,611 and £21,452 per QALY gained for pembrolizumab with axitinib compared to sunitinib, pazopanib, tivozanib and cabozantinib respectively.

4.3.10 Assessment of uncertainty

One-way sensitivity analyses were undertaken and reported in the CS for pairwise comparisons of pembrolizumab and axitinib versus sunitinib and these are presented in a tornado plot. Except for annual discount rates, all other model parameters are varied using the 95% confidence intervals to test the sensitivity of the results to individual parameters or groups of parameters. The results are summarised in the tornado graphs in Figure 6 (CS Figure 37).

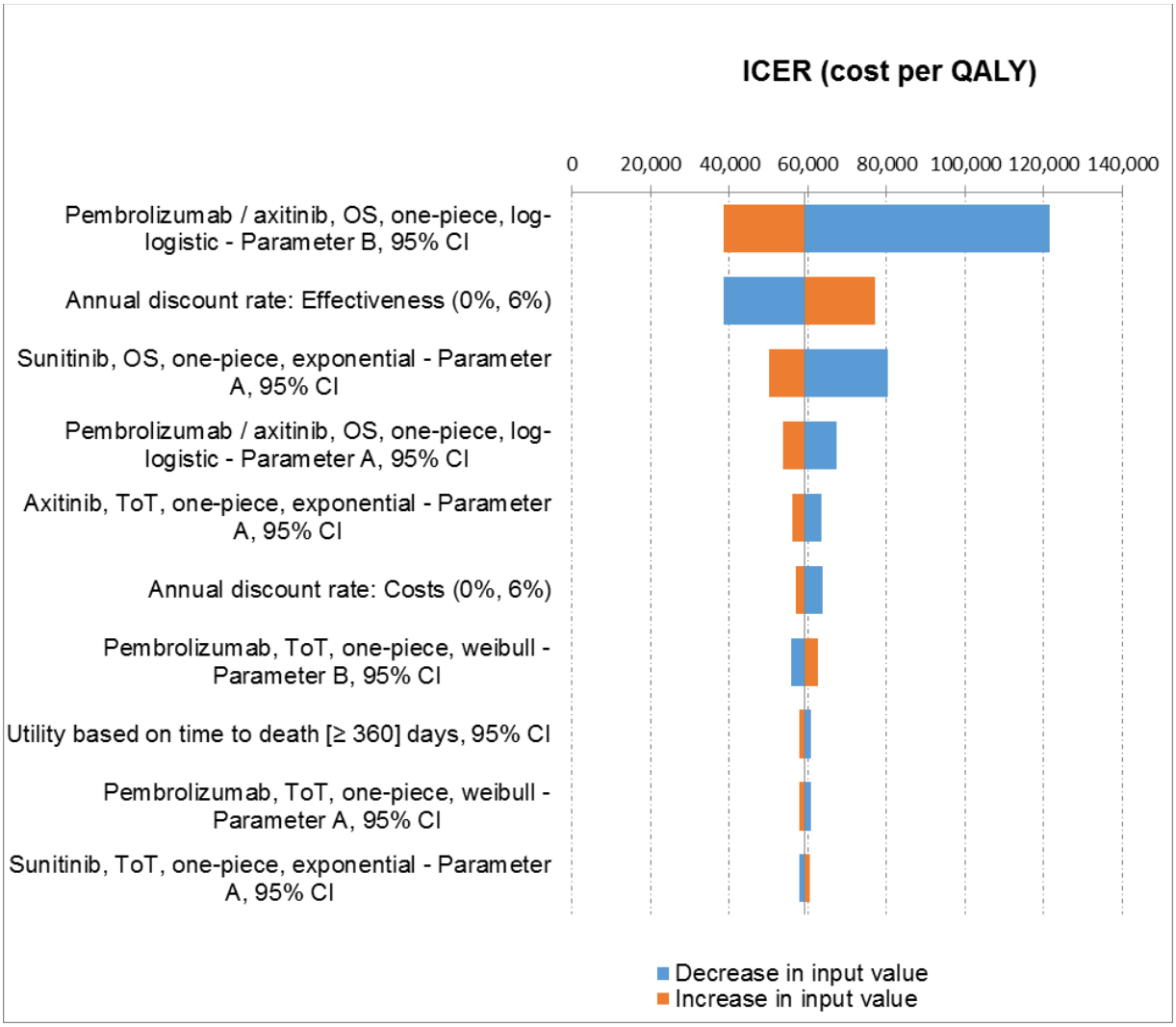

Figure 6 Tornado diagram presenting the results of the deterministic sensitivity analysis for the 20 most influential variables on cost effectiveness results versus sunitinib

Reproduced from CS Figure 37

The company does not justify its method for selecting the parameters reported in the tornado plot or the ranges used for the one-way sensitivity analysis. However, the ERG considers that the use of the 95% CI ranges is reasonable and a well-established way of testing the sensitivity of individual parameters. The parameters of the OS curve for pembrolizumab plus axitinib have the biggest impact on cost-effectiveness, with the ICER increasing by over £50,000 per QALY gained when a parameter of the log-logistic curve is varied. Other significant drivers of cost-effectiveness include annual discount rate for effectiveness and the sunitinib OS curve.

4.3.10.1 Scenario analysis

The company explores a range of scenarios to test structural and methodological uncertainty. These are reported in CS Table 67. It is not clear if the company's scenario analyses were informed by expert opinion. Generally, the company appears to test scenarios using available data that were not used in the base case. We think the parameters explored by the company are reasonable, although we requested additional analyses which were provided in the company's response to our clarification questions (questions B10 to B16). We felt that these additional analyses by the company are incomplete as some of them do not address the questions raised by the ERG. For instance, the ERG requested a scenario analyses for PFS, OS and time on treatment where the same parametric distributions are used for each treatment arm (clarification question B13). However, the company's response does not answer this question. In our base case and scenario analyses, we provide results for these scenarios.

The company found that the biggest source of uncertainty over cost-effectiveness was the introduction of treatment effect waning after 10 years, with an ICER of £86,712 per QALY gained for pembrolizumab with axitinib compared to sunitinib. The choice of OS curve used in the model and the use of alternative modelling approaches for PFS and ToT also increased the ICER significantly. Introducing a 2-year stopping rule for axitinib reduced the ICER to £50,436 per QALY gained.

The company's scenario analyses are shown in Table 42 (CS Table 67).

4.3.10.2 Probabilistic sensitivity analysis

The company's probabilistic sensitivity analysis (PSA) results are summarised in scatterplots, cost effectiveness acceptability curves (CEACs) and in a table of incremental cost per QALY gained (CS Figures 35 and 36: CS Tables 66) for pembrolizumab plus axitinib versus sunitinib. The PSA results, which were estimated for 1000 simulations, are stable and similar to the deterministic results. It takes about 1.5 hours to run a thousand iterations on the company's model. The CEAC and cost effectiveness results are reproduced below in Figure 7 and Table 43 (CS Figure 36 and CS Table 66, respectively).

Table 42 Results from the scenario analyses versus trial comparator sunitinib (list price)

Scenario No.	Description	Pembrolizumab + axitinib vs sunitinib		
		Incremental costs (£)	Incremental QALYs	ICER (£)
Base Case	-	£ 137,537	2.320	£59,292
Scenario 1	Landmark Modelling approach[a]	£ 137,249	2.237	£61,341
Scenario 2	Fully parametric exponential OS extrapolation	£ 135,994	1.861	£73,094
Scenario 3	Fully parametric log-logistic OS extrapolation for pembrolizumab + axitinib, time-constant HR for sunitinib	£ 137,497	2.318	£59,310
Scenario 4	Fully parametric log-logistic OS extrapolation for pembrolizumab + axitinib, time-varying HR for sunitinib	£ 135,616	1.720	£78,854
Scenario 5	Treatment waning after 10 years	£ 134,833	1.555	£86,712
Scenario 6	Alternative modelling approach of PFS and ToT-PFS pembrolizumab + axitinib lognormal; ToT pembrolizumab Weibull, axitinib lognormal.	£ 182,710	2.320	£78,767
Scenario 7	Health state-based utilities (pooled)	£ 137,537	2.169	£63,400
Scenario 8	Health state-based utilities (treatment specific)	£ 137,537	2.259	£60,876
Scenario 9	Removing age-related disutilities	£ 137,537	2.499	£55,045
Scenario 10	Sunitinib dose intensity = 86% (TA169)[71]	£ 133,690	2.320	£57,634
Scenario 11	Removing AE disutilities	£ 137,537	2.319	£59,300
Scenario 12	Trial-based subsequent therapy distribution	£ 141,482	2.320	£60,993
Scenario 13	Axitinib 2 year stopping rule	£ 116,994	2.320	£50,436
Scenario 14	Remove half-cycle correction	£ 137,537	2.320	£59,289

Adapted from CS Table 67

[a] Details shown in CS appendix P (Scenario 1)

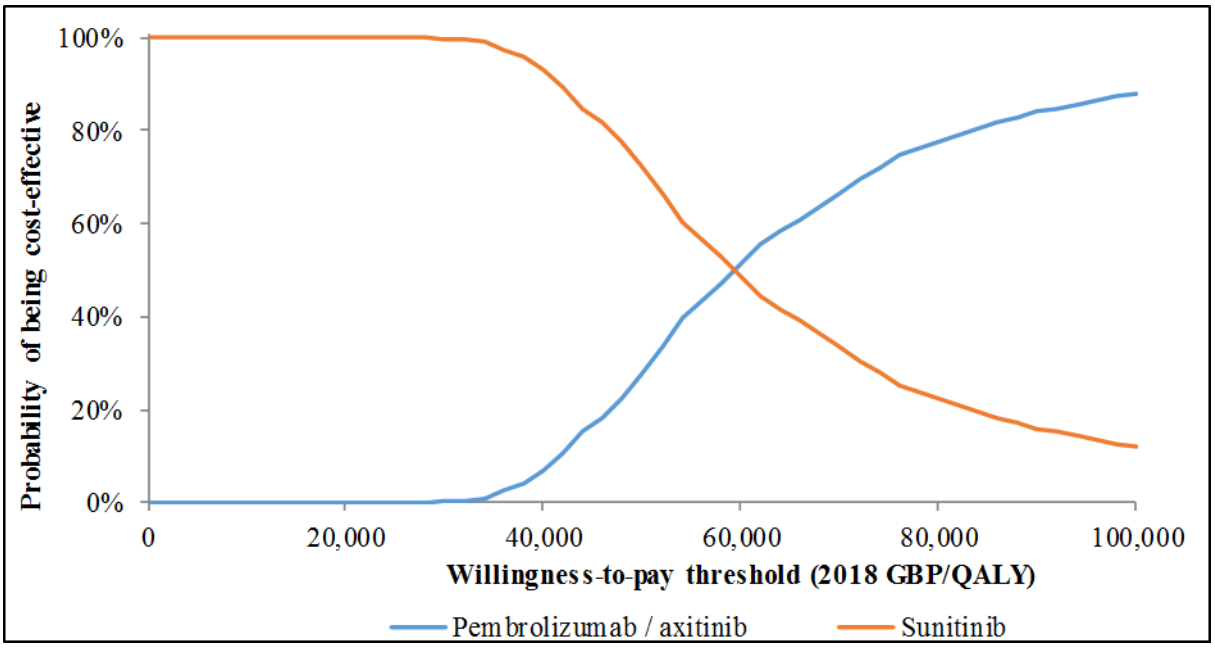

Figure 7 Cost-effectiveness acceptability curve versus sunitinib (list price)

Reproduced from CS Figure 36

Table 43 Incremental cost-effectiveness results based on probabilistic sensitivity analysis versus sunitinib

Intervention	Total Costs	Total QALYs	Incremental Costs	Incremental QALYs	ICER (£/QALY)
Sunitinib	[redacted]	[redacted]	-	-	-
Pembrolizumab + axitinib	[redacted]	[redacted]	£137,352	2.30	£59,726

Reproduced from CS Table 66

The CS reports a 0.3% probability that pembrolizumab plus axitinib is cost-effective at a threshold of £30,000 per QALY gained compared to sunitinib.

All the variables that were included in the PSA are summarised in the CS (Table 62) along with the corresponding distributions. The utility inputs and costs (administrative costs, disease management costs and adverse event management costs) were assigned beta and gamma distributions respectively. We consider these distributions to be appropriate. Drug costs and incidence of AEs for pembrolizumab with axitinib and sunitinib are among parameters not included in the PSA. No justification was provided for the exclusion of these parameters but we consider that drug costs are subject to very little uncertainty, since they are sourced from the BNF.

4.4 Additional work undertaken by the ERG

The ERG did not identify any errors to be corrected in the company model. Table 44 to Table 45 show the assumptions in the company base case and alternatives suggested by the ERG for our base case for the overall patient population. We conduct scenario analyses (Table 46) which use the ERG base case assumptions. We provide justifications for our preferred assumptions. In Table 47, we list the proportion of patients who receive subsequent therapy in our base case and scenario analysis.

Table 44 ERG base case for the extrapolation distributions – overall population

	Overall survival		Progression free survival		Time on treatment		
	Pemb + Ax	Sun	Pemb + Ax	Sun	Pemb	Ax	Sun
Company base case	Log-logistic	Exponential	KM + Exponential	KM + Exponential	Weibull	Exp	Exp

ERG base case	Weibull	Weibull	KM + Exponential	KM + Exponential	Weibull	Weibull	Weibull
Notes	Should use same distribution for both treatments. Weibull provides best fit to sunitinib data.		ERG agrees with company approach.		Should use same distribution for all treatments. Weibull provides best visual fit to data.		

Table 45 ERG base case additional parameters – overall population

Parameter	Company's assumption	ERG preferred assumptions	Reason for ERG preference
Age-adjusted utility	Included age-adjusted utility	Don't include age adjusted utility	See company's response to clarification question B11, no relation found with age and utility.
Subsequent treatment costs	Based on NHS England estimates	ERG base case (see Table 47)	Changes to subsequent treatment: For pembrolizumab + axitinib arm, more patients (20%) would receive cabozantinib See Table 47)
Administration costs	Oral treatments: administration cost of 131.61.	Oral treatments: administration costs £0;	Oral treatments don't normally have costs;
Terminal care cost	£6,789.76	£8,073	Using cost from cabozantinib STA updated to 2017/8.

Table 46 ERG scenarios

Parameter	Company's assumption	ERG preferred assumptions	Scenarios	Reason for ERG preference
Time horizon	40 years	40 years	20 years	20 year horizon used in previous models
Age of cohort	62 years	62 years	57 / 67 years	Exploratory: to assess applicability to the UK RCC population
OS curves	As above	As above	Exponential, Log-logistic for both treatment arms	Other plausible distributions
PFS curves	As above	As above	Weibull, log-logistic	Other plausible distributions
ToT curves	As above	As above	Exponential, log-logistic for both treatment arms	Other plausible distributions
Persistence of OS benefit	No waning effect	No waning effect	Waning effect after 5 / 10 years	Immature OS data. Unclear why there would be a persistence of benefit

				several years after treatment ended.
Time varying HR for PFS and OS	Time varying HR not used	Time varying HR not used	Time varying HR using company's 1st and 2nd / 3rd best fitting FP models	Alternative method to estimate extrapolation of comparator survival curves
Health state utilities	Utilities from company trial for time-to-death	Utilities from company trial for time-to-death	Utilities from previous NICE TAs; tivozanib TA512; pazopanib TA215	
Age-adjusted utility	Included age-adjusted utility	Don't include age adjusted utility	Use age-adjusted utility	See company's response to clarification question B11, no relation found with age and utility.
UK population norms for utility	No adjustment for UK population norms.	No adjustment for UK population norms.	Utility for patients with >360 days to death set to 0.775.	Utility for patients with >360 days to death higher than UK population norms for same age group.
Subsequent treatment costs	Based on NHS England estimates	ERG base case (see Table 47)	ERG scenario analysis (see Table 47)	Based on clinical advice.
	ToT for comparator treatments based on PFS		Apply same assumptions to pembrolizumab / sunitinib	Consistency
Administration costs	Oral treatments: administration cost of 131.61	Oral treatments: administration costs £0;	Oral treatments: administration cost of 131.61;	Oral treatments don't normally have administration costs;

Table 47 ERG base case and scenario analyses on proportion of patients on subsequent-line treatment

	Company base case		**ERG base case**		**ERG scenario analysis**	
Subsequent treatment	**Pembrolizumab + axitinib**	**Sunitinib**	**Pembrolizumab + axitinib**	**Sunitinib**	**Pembrolizumab + axitinib**	**Sunitinib**
Best supportive care	50%	50%	50%	50%	40%	40%
Lenvatinib / everolimus	0%	0%	0%	0%	0%	0%
Axitinib	0%	8%	0%	8%	0%	8%
Cabozantinib	0%	13%	20%	13%	20%	13%
Nivolumab	0%	30%	0%	30%	0%	40%
Pazopanib	30%	0%	20%	0%	25%	0%
Sunitinib	20%	0%	10%	0%	15%	0%

In Table 48, the results for the ERG base-case analysis for the overall population are shown. The pairwise cost-effectiveness results are £120,455, £115,558 and £117,411 per QALY gained for pembrolizumab plus axitinib versus sunitinib, tivozanib and pazopanib respectively. While these results represent our preferred assumptions, they indicate a doubling of the ICERs reported in the CS.

Table 48 ERG base case cost-effectiveness analysis for pembrolizumab + axitinib versus comparators in the overall population (pairwise comparisons)

Treatment	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental cost per QALY gained
Pembrolizumab + axitinib	[redacted]	[redacted]	-	-	-
Sunitinib	[redacted]	[redacted]	£140,895	1.170	£120,455
Tivozanib	[redacted]	[redacted]	£135,168	1.170	£115,558
Pazopanib	[redacted]	[redacted]	£137,335	1.170	£117,411

Table 49 shows the scenario analyses and the effect of these on model results. The results vary between £72,591 - £162,424 per QALY gained for pembrolizumab plus axitinib compared to sunitinib. The scenario analyses are described in more detail in Table 46. Those scenarios which have a large effect on model results are changes to the distributions used for OS, using the log-logistic curve for ToT, including a waning effect and changes to the utility values.

Table 49 ERG scenario analyses for pembrolizumab + axitinib versus sunitinib in the overall population

Scenario	Scenarios	Incremental costs	Incremental QALYs	ICER (£/QALY)
Base case		£140,895	1.170	£120,455
Time horizon	20 years	£140,779	1.149	£122,498
Age of cohort	57 years	£140,895	1.170	£120,447
	67 years	£140,894	1.169	£120,510
OS curves	Exponential	£143,209	1.973	£72,591
	Log-logistic	£141,615	1.419	£99,790
PFS curves	Weibull	£140,996	1.170	£120,541
	Log-logistic	£141,019	1.170	£120,561
ToT curves	Exponential	£141,627	1.170	£121,080
	Log-logistic	£166,512	1.170	£142,356
Persistence of OS benefit	Waning effect after 5 years	£137,625	0.847	£162,424
	Waning effect after 10 years	£140,534	1.086	£129,368

Time varying HR for PFS and OS	Company best fitting FP model	£140,784	1.162	£121,183
	Company 2nd best fitting FP model[a]	£140,569	1.074	£130,897
Health state utilities	Utilities from Tivozanib TA512;	£140,895	0.953	£147,873
	Utilities from pazopanib TA215	£140,895	0.883	£159,484
Population norms utility	Utility set at 0.775 for time to death > 360 days	£140,895	1.100	£128,044
Age-adjusted utility	Use age-adjusted utility	£140,895	1.124	£125,389
Subsequent treatment costs	ERG scenario analysis (see Table 47)	£138,591	1.170	£118,485
Administration costs	Oral treatments: administration cost of £131.61;	£140,527	1.170	£120,140

[a] fractional polynomial NMA 2nd best fitting model (company clarification response document appendix Table 43, 44).

Subgroup analysis: intermediate / poor risk group

We also conducted analyses for the intermediate / poor risk group using the ERG's preferred base case assumptions for the overall patient population (Table 44 and Table 45). The results of these are shown in Table 50. The ICER for pembrolizumab plus axitinib compared to cabozantinib is £48,424 per QALY gained.

Table 50 ERG analysis of cost-effectiveness for pembrolizumab + axitinib versus comparators in the intermediate / poor risk subgroup (Pairwise comparisons)

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental cost per QALY gained
Pembrolizumab + axitinib	[redacted]	[redacted]	-	-	-
Sunitinib	[redacted]	[redacted]	£141,941	1.010	£140,481
Tivozanib	[redacted]	[redacted]	£137,480	1.010	£136,065
Pazopanib	[redacted]	[redacted]	£139,200	1.010	£137,768
Cabozantinib	[redacted]	[redacted]	£44,012	0.909	£48,424

Table 51 shows the ERG scenario analyses and the effect of these on the model results for the intermediate / poor subgroup. The results vary between £27,892 - £149,347 per QALY gained for pembrolizumab plus axitinib compared to cabozantinib. The scenario analyses are described in more detail in Table 46. Those scenarios which have a large effect on model results are

changes to the distributions used for PFS, using the log-logistic curve for ToT, including a treatment effect waning assumption, using time varying hazards from the fractional polynomial NMA and changes to the utility values.

For the two time varying hazard fractional polynomial models used, the ICER varies between £117,279 and £149,347 per QALY gained. The ERG's critique of the fractional polynomial approach is given in section 3.1.7.5. In this section we note that the company prefers to use the constant HR NMA because the results are more stable than the results of the time varying hazards NMA for the intermediate / poor risk subgroup. The results from our analyses show a large variability in cost-effectiveness compared to the base case which confirms the instability of this approach.

The ERG results shown in Table 48 - Table 51 are calculated using the list price for all treatments. We submitted to NICE a separate confidential appendix which uses the confidential discount prices agreed with the NHS for all treatments for the company and ERG base case analyses.

Table 51 Scenario analyses for pembrolizumab + axitinib versus cabozantinib in the intermediate / poor risk population

Scenario	Scenarios	Incremental costs	Incremental QALYs	ICER (£/QALY)
ERG base case		£44,012	0.909	£48,424
Time horizon	20 years	£43,989	0.904	£48,645
Age of cohort	57 years	£44,012	0.909	£48,424
	67 years	£44,011	0.909	£48,425
OS curves	Exponential	£46,146	1.265	£36,489
	Log-logistic	£46,040	1.651	£27,892
PFS curves	Weibull	£59,261	0.909	£65,201
	Log-logistic			Implausible
ToT curves	Exponential	£40,397	0.909	£44,447
	Log-logistic	£83,907	0.909	£92,318
Persistence of OS benefit	Waning effect after 5 years	£50,525	0.689	£73,290
	Waning effect after 10 years	£44,651	0.872	£51,223
	Best-fitting FP model	£38,473	0.258	£149,347

Time varying HR for PFS and OS	3rd best-fitting FP model[a]	£42,805	0.365	£117,279
Health state utilities	Utilities from tivozanib TA512;	£44,012	0.673	£65,401
	Utilities from pazopanib TA215	£44,012	0.591	£74,530
Population norms utility	Utility set at 0.775 for time to death > 360 days	£44,012	0.855	£51,469
Age-adjusted utility	Use age-adjusted utility	£44,012	0.878	£50,108
Subsequent treatment costs	ERG scenario analysis (see Table 47)	£45,862	0.909	£50,460
Administration costs	Oral treatments: administration cost of 131.61.	£41,639	0.909	£45,813

[b] OS only, company clarification question response appendix Table 129, Table 130. PFS uses constant HR

5 End of life

The CS does not consider pembrolizumab plus axitinib to meet the NICE end of life criteria for the overall RCC patient population (CS Table 39). Estimates of OS for sunitinib in pivotal phase III RCTs are in excess of 24 months (criterion 1 states that the treatment is indicated for patients with a short life expectancy, normally less than 24 months). The ERG agrees with this assertion. However, the CS claims that patients in the IMDC poor risk sub-group would meet the end of life criteria as they have a life expectancy of less than 24 months, and an expected increase in life expectancy of greater than three months with pembrolizumab plus axitinib. The CS appears to use sunitinib as the standard of care for estimating life expectancy and gains in life years in this patient subgroup. However, the ERG notes that cabozantinib is specifically recommended by NICE in poor (and intermediate) risk patients, based on NICE TA542[24] Of note, the CS does not explicitly state the rationale for the choice of the poor risk subgroup when in their assessment of clinical effectiveness and cost effectiveness the subgroup is intermediate / poor risk. Thus, it is not possible for the ERG to generate modelled estimates of OS for poor risk subgroup patients to inform end of life assessment.

In Table 52 we summarise and critique the company's evidence in support of their case for end of life criteria applying to poor risk RCC patients.

Table 52 Summary and critique of the CS case for meeting end of life criteria in poor risk RCC patients

Criterion	Data available	ERG comment
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	The CS cites pivotal phase III trials of first line RCC treatments, including CABOSUN, which included intermediate / poor RCC risk patients. Median OS was 30.3 months for cabozantinib, and 21.8 months for sunitinib. Other trial estimates of OS for sunitinib were in excess of 24 months (though not restricted to intermediate / poor risk patients). The CS cites final results from extended follow-up of a global, expanded-access trial of sunitinib treatment in 4543 patients with metastatic RCC ineligible for registration trials.[72] Median OS stratified by risk group was 56.5 months (favourable risk), 20.0 months (intermediate risk), and 9.1 months (poor risk). The distribution of patients across IMDC risk categories was 22%, 48% and 20%, respectively.	The median OS of 30.3 months for intermediate / poor risk patients in the CABOSUN trial exceeds the end of life criterion of less than 24 months life expectancy. This is a large study reflective of a real world population. However, cabozantinib is not included in this study, which is one of the NICE recommended treatment options for patients at intermediate / poor risk.
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	The CS provides median OS rates for pembrolizumab plus axitinib versus sunitinib from KEYNOTE-426, at 12 months. The CS also provides OS rates from their economic model at 2 years and 3 years. The ERG notes that these are for the overall RCC population, rather than the poor risk population.	The ERG reports mean undiscounted life years based on the company's model, and the ERG's modelled base case (Table 53), for the intermediate / poor risk subgroup. Pembrolizumab + axitinib extended life by greater than 3 months compared to sunitinib and

	The CS does not attempt to translate these OS rates into life years gained.	cabozantinib, in both the ERG and the company's base case models.

Table 53 ERG and company modelled estimates of overall survival in the intermediate / poor risk subgroup

Treatment	Mean undiscounted life years	
	ERG base case modelled estimate	Company base case modelled estimate
Pembrolizumab + axitinib	4.492	7.691
Sunitinib	3.000	3.266
Cabozantinib	3.129	4.664

ERG conclusion

The ERG agrees with the company that pembrolizumab plus axitinib does not meet the first end of life criterion in the overall RCC population (treatment is indicated in patients with a short life expectancy, normally less than 24 months). The ERG disagrees with the company that pembrolizumab plus axitinib meets the first end of life criterion in the poor RCC risk subgroup, based on cabozantinib being specifically recommended by NICE in this subgroup. The ERG is in agreement with the company that pembrolizumab plus axitinib meets the second end of life criterion (treatment offers an extension to life, normally of at least an additional three months, compared with current NHS treatment). We are therefore of the opinion that pembrolizumab plus axitinib 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 company considers pembrolizumab in itself to be innovative in the first line treatment of RCC, noting that it is available for a wide range of indications. It has a Breakthrough Therapy Designation by the US Food and Drug Administration and a positive scientific opinion from the UK MHRA's Early Access schemes for some of these indications. The company also considers that the innovative immuno-oncology combination regimen of pembrolizumab plus axitinib represents a "step-change" in the management of RCC (CS page 79) as it targets both

angiogenesis and immune-checkpoint pathways. The CS states that other novel anticancer agents have shown improvements over the original immunotherapies, but there remains an unmet need because disease progression occurs in most people within two years. The pembrolizumab plus axitinib combination provides additional clinical benefit over the standard of care. The company states that pembrolizumab should be considered innovative by its potential to make a significant and substantial impact in an area of high unmet need. The ERG clinical advisors agree there does remain an element of unmet need and that the rationale for the combination in RCC is made, however there are other potential treatments that should be considered in relation to pembrolizumab and axitinib (e.g. avelumab plus axitinib – currently the subject of a separate NICE technology appraisal).[73 , 74]

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The company's decision problem is largely consistent with the NICE scope, although the population in the CS is restricted to patients with clear cell RCC. The results will not be generalisable to patients with non-small cell RCC types (approximately 25% of patients). The ERG notes that previous NICE appraisals of treatments for RCC also did not restrict the scope to clear cell, despite the pivotal trials comprising mostly or exclusively of clear cell RCC patients. The current CS is therefore in line with evidence accepted in previous NICE appraisals.

The evidence for clinical effectiveness of pembrolizumab plus axitinib is from a large multinational RCT, KEYNOTE-426. The outcomes and statistical analyses of the trial are appropriate, and other than its open-label design, the trial has a low risk of bias. The generalisability of the trial to the UK population is uncertain, as most participants were randomised outside of Europe and less than 6% were from the UK. The participants in the trial are younger and fitter than a typical population with advanced untreated RCC, but similar in these aspects to other pivotal trials of treatments in this indication appraised by NICE.

At the first interim analysis, KEYNOTE-426 demonstrated a significant improvement in PFS with pembrolizumab plus axitinib (15.1 months) compared with sunitinib (11.1 months). Median OS was not reached in either arm. Efficacy testing was stopped early at the first interim analysis,

which can sometimes result in over-estimation of treatment effect. In this case the ERG considers it is unlikely that PFS has been over-estimated, but OS results should be viewed with caution as they are immature.

The company conducted an NMA in order to indirectly compare pembrolizumab plus axitinib with the other treatments in the NICE scope (tivozanib, pazopanib, and cabozantinib for intermediate/poor risk according to IMDC criteria). Previous NICE appraisals of first line treatments for advanced RCC have accepted the assumption that sunitinib, pazopanib and tivozanib are broadly similar to each other in efficacy, and therefore the committees have not considered indirect comparisons as a key factor in their decision making. In the current appraisal the company likewise assumes that pazopanib and tivozanib are similar to sunitinib, and therefore use the direct comparison between pembrolizumab plus axitinib from the KEYNOTE-426 trial to inform clinical effectiveness estimates in the model (i.e. the NMA is not used in the model). However, there is no direct trial comparison between pembrolizumab plus axitinib and cabozantinib, the comparator treatment relevant to patients in the intermediate / poor risk patient subgroup. Therefore, the indirect comparison of these two treatment regimens via NMA is of importance as it informs the economic model cost effectiveness estimates for this subgroup. Overall, the ERG considers the methods and assumptions used to conduct the NMAs to have been appropriately exercised, though the results of the intermediate / poor risk sub-group NMA should be treated with caution as it is based on a sub-set of the randomised population of the KEYNOTE-426 trial, rather than the full trial population.

7.2 Summary of cost effectiveness issues

The company's base case analysis of pembrolizumab plus axitinib versus sunitinib, based on extrapolation curves for OS, PFS and TTD from the overall population of the KEYNOTE–426 trial, gave an ICER of £59,292 per QALY gained. Pazopanib and tivozanib were considered clinically equivalent to sunitinib. In the company's analysis, pazopanib and tivozanib were slightly more expensive than sunitinib and when compared with pembrolizumab plus axitinib, there was an ICER of £57,540 and £56,648 per QALY gained for tivozanib and pazopanib.

The company also provided a comparison against cabozantinib in the intermediate / poor risk RCC population (as defined by the IMDC criteria). The ICER was £21,452 per QALY gained for pembrolizumab plus axitinib compared to cabozantinib.

The ERG identified a number of uncertainties in the company's model and tested an alternative set of assumptions and input parameters relating to the method of fitting the OS and TTD curves, age-adjusted utility, administration costs and terminal care costs.

The ERG-preferred analyses gave higher ICER estimates: £120,455 per QALY for pembrolizumab plus axitinib compared with sunitinib for the overall population and an estimated ICER of £48,424 per QALY for pembrolizumab plus axitinib compared with cabozantinib in the intermediate / poor risk subgroup.

The above analyses have been completed using list price for the treatments. We present results for the above analyses using existing PAS discounts for first and subsequent line treatments in a confidential addendum to this report.

8 REFERENCES

1. Cancer Research UK. Kidney cancer: stages, types and grades. Secondary Kidney cancer: stages, types and grades. https://www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/types-grades.
2. Kidney Care UK. What is Kidney Cancer. Secondary What is Kidney Cancer. https://www.kcuk.org.uk/kidneycancer/what-is-kidney-cancer/.
3. NHS. Overview: Kidney Cancer. Secondary Overview: Kidney Cancer. 2016. https://www.nhs.uk/conditions/kidney-cancer/.
4. Kidney Care UK. Kidney Cancer. Secondary Kidney Cancer. https://www.kidneycareuk.org/about-kidney-health/conditions/kidney-cancer/.
5. Cancer Research UK. Kidney cancer statistics. Secondary Kidney cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer#heading-Zero.
6. Harding G, Cella D, Robinson DJ, et al. Symptom burden among patients with Renal Cell Carcinoma (RCC): content for a symptom index. Health and Quality of Life Outcomes 2007;**5**.
7. Cella D. Beyond Traditional Outcomes: Improving Quality of Life in Patients with Renal Cell Carcinoma. The Oncologist 2011;**16**:23-31.
8. Albiges L, Powles T, Staehler M, et al. Updated European Association of Urology Guidelines on Renal Cell Carcinoma: Immune Checkpoint Inhibition Is the New Backbone in First-line Treatment of Metastatic Clear-cell Renal Cell Carcinoma. European Urology 2019;**76**:151-56.
9. Heng D, Xie W, Regan M, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor–targeted agents: results from a large, multicenter study. Journal of Clinical Oncology 2009;**34**:5794-99.
10. Heng D, Xie W, Regan M, et al. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. . Lancet Oncology 2013;**14**:141-48.

11. Choueiri K, Halabi S, Sanford B, 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 2017;**35**.
12. Motzer R, Hutson T, Cella D, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. . New England Journal of Medicine 2013;**369**:722-31.
13. Motzer R, Nosov D, Eisen T, et al. Tivozanib Versus Sorafenib As Initial Targeted Therapy for Patients With Metastatic Renal Cell Carcinoma: Results From a Phase III Trial. Journal of Clinical Oncology 2013;**31**:3791-99.
14. Sternberg CN, Motzer RJ, Hutson TE, et al. COMPARZ Post Hoc Analysis: Characterizing Pazopanib Responders with Advanced Renal Cell Carcinoma. Clinical Genitourinary Cancer 2019.
15. George DJ, Hessel C, Halabi S, et al. Cabozantinib Versus Sunitinib for Untreated Patients with Advanced Renal Cell Carcinoma of Intermediate or Poor Risk: Subgroup Analysis of the Alliance A031203 CABOSUN trial. Oncologist 2019.
16. Rini B, Plimack E, Stus V, et al. Pembrolizumab plus Axitinib versus Sunitinib for Advanced Renal-Cell Carcinoma. New England Journal of Medicine 2018;**380**:1116-27.
17. Bedard G, Zeng L, Zhang L, et al. Minimal important differences in the EORTC QLQ-C30 in patients with advanced cancer. Asia-Pacific Journal of Clinical Oncology 2014;**10**(2):109-17.
18. Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. New England Journal of Medicine 2018;**378**(14):1277-90.
19. Guyatt G, Briel M, Glasziou P, et al. Problems of stopping trials early. BMJ 2012;**344**:e3863.
20. Freidlin B, Korn E. Stopping clinical trials early for benefit: impact on estimation. Clinical Trials 2009;**6**:119 - 25.
21. Mehta A, Sonpavde G, Escudier B. Tivozanib for the treatment of renal cell carcinoma: results and implications of the TIVO-1 trial. Future oncology 2014;**10**:1819-26.

22. Escudier B, Szczylik C, Hutson T, et al. Randomized phase II trial of first-line treatment with sorafenib versus interferon Alfa-2a in patients with metastatic renal cell carcinoma. Journal of clinical oncology 2009;**27**:1280-89.
23. Motzer R, Hutson T, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. New England Journal of Medicine 2007;**356**:115-24.
24. National Institute for Health and Care Excellence. TA542 Cabozantinib for untreated advanced renal cell carcinoma. 2018.
25. Dias S, Welton N, Sutton A, et al. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials (last updated September 2016). Secondary NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials (last updated September 2016) 2011. http://www.nicedsu.org.uk.
26. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC Medical Research Methodology 2011;**11**(1):61.
27. National Institute for Health and Care Excellence. TA512: Tivozanib for treating advanced renal cell carcinoma. 2018.
28. Jansen J, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011;**14**:417-28.
29. Turner RM, Jackson D, Wei Y, et al. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Statistics in Medicine 2015;**34**(6):984-98.
30. Zondervan-Zwijnenburg M, Peeters M, Depaoli S, et al. Where Do Priors Come From? Applying Guidelines to Construct Informative Priors in Small Sample Research. Research in Human Development 2017;**14**(4):305-20.
31. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and Quality of Life Outcomes 2007;**5**(70):http://www.hqlo.com/content/5/1/70.

32. Liu G, Wang J, Liu K, et al. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Statistics in Medicine 2006;**25**:1275–86.
33. Atkins M, Plimack E, Puzanov I, et al. Axitinib in combination with pembrolizumab in patients with advanced renal cell cancer: a non-randomised, open-label, dose-finding, and dose-expansion phase 1b trial. Lancet Oncology 2018;**19**:405-15.
34. Amdahl J, Diaz J, Sharma A, et al. Cost-effectiveness of pazopanib versus sunitinib for metastatic renal cell carcinoma in the United Kingdom. PloS one 2017;**12**(6):e0175920.
35. Hoyle M, Green C, Thompson-Coon J, et al. Cost-Effectiveness of Temsirolimus for First Line Treatment of Advanced Renal Cell Carcinoma. Value in Health 2010;**13**(1):61-68.
36. Kilonzo M, Hislop J, Elders A, et al. Pazopanib for the first-line treatment of patients with advanced and/or metastatic renal cell carcinoma. Pharmacoeconomics 2013;**31**(1):15-24.
37. National Institute for Health and Care Excellence. TA169 Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma. 2009. https://www.nice.org.uk/guidance/ta169/documents/ta169-renal-cell-carcinoma-sunitinib-review-decision-may-2012.
38. Scottish Medicines Consortium. Sunitinib (Sutent) First line treatment of advanced and/or metastatic renal cell carcinoma (MRCC). 2007.
39. Scottish Medicines Consortium. Pazopanib (Votrient) For the first-line treatment of advanced renal cell carcinoma (RCC) and for patients who have received prior cytokine therapy for advanced disease. 2011.
40. Thompson-Coon J, Hoyle M, Green C, et al. Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. 2010.
41. National Institute for Health and Care Excellence. TA581 Nivolumab with ipilimumab for untreated advanced renal cell carcinoma. 2019.

42. Porta C. Regarding:'Costs of managing adverse events in the treatment of first-line metastatic renal cell carcinoma: bevacizumab in combination with interferon-α2a compared with sunitinib'. British journal of cancer 2010;**102**(7):1196.
43. Chen J, Hu G, Chen Z, et al. Cost-effectiveness Analysis of Pembrolizumab Plus Axitinib Versus Sunitinib in First-line Advanced Renal Cell Carcinoma in China. Clinical Drug Investigation 2019:1-8.
44. National Institute for Health and Care Excellence. TA215: Pazopanib for the first-line treatment of advanced renal cell carcinoma. 2011.
45. Latimer N. Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. NICE DSU Technical Document 14 2011.
46. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. NICE, London 2013.
47. Merck Sharp & Dohme Corp. asoMC. Protocol of A Phase III Randomized, Open-label Study to Evaluate Efficacy and Safety of Pembrolizumab (MK-3475) in Combination with Axitinib versus Sunitinib Monotherapy as a First-line Treatment for Locally Advanced or Metastatic Renal Cell Carcinoma (mRCC) (KEYNOTE-426), 2018.
48. National Institute for Health and Care Excellence. TA428. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy. 2017.
49. National Institute for Health and Care Excellence. TA366. Pembrolizumab for advanced melanoma not previously treated with ipilimumab. 2017.
50. Hatswell AJ, Pennington B, Pericleous L, et al. Patient-reported utilities in advanced or metastatic melanoma, including analysis of utilities by time to death. Health Qual Life Outcomes 2014;**12**:140.
51. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. Discussion paper 172 1999.
52. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health 2010;**13**(5):509-18.
53. Ara R, Wailoo AJ. NICE DSU Technical Support Document 12: The use of health state utility values in decision models. 2011.

54. Joint Formulary Committee. *British National Formulary*: London: British Medical Association and Royal Pharmaceutical Society of Great Britain, 2019.
55. Motzer RJ, Hutson TE, Glen H, et al. Lenvatinib, everolimus, and the combination in patients with metastatic renal cell carcinoma: a randomised, phase 2, open-label, multicentre trial. The lancet oncology 2015;**16**(15):1473-82.
56. National Institute for Health and Care Excellence. TA333: Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment. 2015.
57. National Institute for Health and Care Excellence. TA417: Nivolumab for previously treated advanced renal cell carcinoma. 2016.
58. Motzer RJ, Escudier B, Tomczak P, et al. Axitinib versus sorafenib as second-line treatment for advanced renal cell carcinoma: overall survival analysis and updated results from a randomised phase 3 trial. The Lancet Oncology 2013;**14**(6):552-62.
59. National Institute for Health and Care Excellence. TA463: Cabozantinib for previously treated advanced renal cell carcinoma. 2017.
60. Motzer RJ, Escudier B, Powles T, et al. Long-term follow-up of overall survival for cabozantinib versus everolimus in advanced renal cell carcinoma. British journal of cancer 2018;**118**(9):1176.
61. Sternberg CN, Hawkins RE, Wagstaff J, et al. A randomised, double-blind phase III study of pazopanib in patients with advanced and/or metastatic renal cell carcinoma: final overall survival results and safety update. European journal of cancer 2013;**49**(6):1287-96.
62. National Institute for Health and Care Excellence. TA432: Everolimus for advanced renal cell carcinoma after previous treatment. 2017. https://www.nice.org.uk/guidance/ta432.
63. Hutson TE, Escudier B, Esteban E, et al. Randomized phase III trial of temsirolimus versus sorafenib as second-line therapy after sunitinib in patients with metastatic renal cell carcinoma. Journal of Clinical Oncology 2014;**32**(8):760.
64. 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.

65. Rini BI, Halabi S, Rosenberg JE, et al. Phase III trial of bevacizumab plus interferon alfa versus interferon alfa monotherapy in patients with metastatic renal cell carcinoma: final results of CALGB 90206. Journal of clinical oncology 2010;**28**(13):2137.
66. Rini BI, Plimack ER, Stus V, et al. Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. New England Journal of Medicine 2019;**380**(12):1116-27.
67. Merck Sharp & Dohme Corp. A Phase III Randomized, Open-label Study to Evaluate Efficacy and Safety of Pembrolizumab (MK-3475) in Combination with Axitinib versus Sunitinib Monotherapy as a First-line Treatment for Locally Advanced or Metastatic Renal Cell Carcinoma (mRCC) (KEYNOTE-426), 2018.
68. Department of Health. NHS reference costs 2017-2018. http://wwwdhgovuk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_122803 2018.
69. Curtis L, Burns A. Unit Costs of Health and Social Care 2018. http://wwwpssruacuk/pdf/uc/uc2010/uc2010pdf 2019.
70. National Institute for Health and Care Excellence. TA519 Pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy. 2018.
71. NICE. TA169: Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma. NICE website: NICE, 2009.
72. Gore ME, Szczylik C, Porta C, et al. Final results from the large sunitinib global expanded-access trial in metastatic renal cell carcinoma. Br J Cancer 2015;**113**(1):12-9.
73. National Institute for Health and Care Excellence. Avelumab with axitinib for untreated advanced or metastatic renal cell carcinoma [ID1547] 2019. (accessed 03/09/2019).
74. Motzer R, Penkov K, Haanen J, et al. Avelumab plus Axitinib versus Sunitinib for Advanced Renal-Cell Carcinoma. New England Journal of Medicine 2019;**380**:1103-15.

75. Choueiri TK, Hessel C, Halabi S, et al. Cabozantinib versus sunitinib as initial therapy for metastatic renal cell carcinoma of intermediate or poor risk (Alliance A031203 CABOSUN randomised trial): Progression-free survival by independent review and overall survival update. European journal of cancer (Oxford, England : 1990) 2018;**94**:115-25.

9 APPENDICES

9.1 NICE appraisal committee conclusions on equivalence of treatment comparisons in previous appraisals of treatments for first line advanced RCC.

NICE TA215 Pazopanib for the first-line treatment of advanced renal cell carcinoma[44]

"The Committee concluded that pazopanib is likely to be more clinically effective than interferon-α and is probably comparable in its effectiveness to sunitinib. Subsequent publication of the COMPARZ trial in which sunitinib and pazopanib were directly compared confirmed this assertion, though the safety profile and HRQoL was better for patients treated with pazopanib.

NICE TA512 Tivozanib for treating advanced renal cell carcinoma[27]

"The committee concluded that it had seen no evidence to suggest that tivozanib was more effective than sunitinib or pazopanib in extending overall and progression-free survival. What evidence there was suggested that, at best, tivozanib may have a similar effect to sunitinib or pazopanib".

NICE TA542 Cabozantinib for untreated advanced renal cell carcinoma[24]

"The committee recalled that pazopanib and sunitinib can be considered equally clinically effective. Therefore, it concluded that an indirect treatment comparison was not needed, and did not consider it further".

NICE TA581 Nivolumab with ipilimumab for untreated advanced renal cell carcinoma[41]

"The committee recalled that pazopanib and sunitinib can be considered equally clinically effective. It concluded that an indirect treatment comparison was not needed and did not consider it further".

(NB. Nivolumab with ipilimumab is not a comparator in this current appraisal).

9.2 ERG critical appraisal of relevant comparator treatment trials included in network meta-analysis

Critical appraisal of the CABOSUN trial[11]

NICE quality assessment criteria for RCT	Judgement
1. Was the method used to generate random allocations adequate?	Unclear risk of bias
Comments: Stratified randomisation using a dynamic allocation method to balance prognostic factors between treatment groups, no further details.	
2. Was the allocation adequately concealed?	Unclear risk of bias
Comments: The method of allocation concealment is not reported in the trial publication or study protocol	
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Yes (low risk of bias)
Comments: The publication states that overall, the treatment groups were balanced with respect to baseline demographic and disease characteristics.	
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 (high risk of bias)
Comments: Open label trial.	
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	Unclear risk of bias
Comments: drop out balanced for withdrawal due to progression and AEs but there were differences between the study arms in the number of patients who did not receive the study drug and in the number of patients who withdrew consent.	
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	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 (low risk of bias) Yes Yes
Comments: ITT approach (all patients who were randomised) for all but safety data (the safety analysis population was patients who received ≥1 dose of study drug).	

Critical appraisal of the COMPARZ trial[12]

NICE quality assessment criteria for RCT	Judgement
1. Was the method used to generate random allocations adequate?	Unclear risk of bias
Comments: States patients were randomly assigned to one of the two study drugs in a 1:1 ratio in permuted blocks of four but method used to generate the schedule not reported.	
2. Was the allocation adequately concealed?	Yes (low risk of bias)
Comments: Interactive voice response system used.	

3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Yes (low risk of bias)
Comments: No notable differences between the groups in demographic or clinical characteristics	
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 (high risk of bias)
Comments: The trial was open-label. Imaging data were re-evaluated by an independent review committee who 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 (low risk of bias)
Comments: The number of treatment discontinuations was similar between the two groups	
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No (low risk of bias)
Comments: Outcome data are reported for each of the stated 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 (low risk of bias) Yes Unclear
Comments: Efficacy data were analysed in the ITT 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.	

Critical appraisal of the TIVO-1 trial[13 , 21]

NICE quality assessment criteria for RCT	**Judgement**
1. Was the method used to generate random allocations adequate?	Unclear risk of bias
Comments: States that randomisation was stratified (geographical region, number of prior treatments for metastatic disease, number of metastatic sites/ organs) but no details of the method to generate the sequence	
2. Was the allocation adequately concealed?	Unclear risk of bias
Comments: Not reported	
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Unclear risk of bias
Comments: Study reports some imbalance between groups for ECOG performance status 0 or 1 which may be prognostic.	
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 (High risk of bias)
Comments: open label trial, response and progression outcomes were evaluated by a blinded independent radiology reviewer but other outcomes were not assessed blind.	

5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	Unclear risk of bias
Comments: Numbers discontinuing treatment differed but no details of numbers discontinuing the study were reported.	
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No (Low risk of bias)
Comments: All outcomes stated in the methods are reported	
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 (Low risk of bias) Yes Unclear
Comments: No details reported	

9.3 Differences in source data and results of constant hazards NMA: CS vs ERG analysis

Differences in NMA source data between CS and ERG

		Data as reported in CS Tables 22, 24, 26, 28			Data as reported in trial publications (extracted by ERG). KEYNOTE-426 taken from CS					
Trial ID	**Comparison**	**HR**	**LHR**	**LSE**	**HR**	**LCI**	**UCI**	**LHR**	**LSE**	**Notes**
Base case PFS										
COMPARZ	pazopanib vs sunitinib	1.05	0.05	0.08	1.05	0.9	1.22	0.05	0.08	
Escudier et al 2009	sorafenib vs INFα	1.14	0.13	0.19	1.14	0.79	1.64	0.13	0.22	calculated reciprocal
	INFα vs sorafenib				0.88	0.61	1.27	-0.13	0.17	Pre-crossover data (Period 1)
KEYNOTE-426	Pembrolizumab+axitinib vs sunitinib	0.69	-0.37	0.1				-0.37	0.1	
Motzer et al 2007	INFα vs sunitinib	1.86	0.62	0.09	2.38	1.85	3.13	0.87	0.32	calculated reciprocal
	sunitinib vs INFα				0.42	0.32	0.54	-0.87	0.06	
TIVO-1	tivozanib vs sorafenib	0.76	-0.28	0.14	0.756	0.58	0.985	-0.28	0.10	Treatment naïve subgroup
Base case OS										
COMPARZ	pazopanib vs sunitinib	0.92	-0.08	0.07	0.91	0.76	1.08	-0.09	0.08	
KEYNOTE-426	Pembrolizumab+axitinib vs sunitinib	0.53	-0.63	0.17				-0.63	0.17	
Intermediate / poor risk subgroup PFS										
CABOSUN	cabozantinib vs sunitinib	0.48	-0.73	0.22	0.66	0.46	0.95	-0.42	0.13	CS used independent committee PFS;[75] ERG used investigator PFS (primary outcome)[11]
KEYNOTE-426	Pembrolizumab+axitinib vs sunitinib	0.67	-0.4	0.12				-0.4	0.12	
Intermediate / poor risk subgroup OS										
CABOSUN	cabozantinib vs sunitinib	0.8	-0.22	0.21	0.8	0.53	1.21	-0.22	0.17	Updated paper (Choueiri, 2018)

KEYNOTE-426	Pembrolizumab+axitinib vs sunitinib	0.52	-0.65	0.18				-0.65	0.18	

HR = Hazard ratio; LHR = log hazard ratio; LSE = log standard error; LCI = lower confidence interval; UCI = upper confidence interval; Shaded cells indicate disagreement between CS and ERG data estimates

Comparison of CS and ERG results: constant HRs (vs sunitinib)

	CS Tables 23,25,27,29)		ERG scenario	
	HR	**95% CrI**	**HR**	**95% CrI**
Base case PFS				
pazopanib	1.05	0.90,1.23	1.05	0.90, 1.23
sorafenib	2.11	1.40,3.18	2.74	1.92, 3.81
INFα	1.85	1.55, 2.22	2.38	2.13,2.66
Pembrolizumab+axitinib	0.69	0.57,0.84	0.69	0.57,0.84
tivozanib	1.6	0.98, 2.59	2.08	1.37, 3.05
Base case OS				
pazopanib	0.92	0.80, 1.07	0.91	0.78, 1.07
Pembrolizumab+axitinib	0.53	0.38, 0.74	0.54	0.38, 0.74
Intermediate / poor risk subgroup PFS				
cabozantinib	0.48	0.31, 0.74	0.67	0.52,0.84
Pembrolizumab+axitinib	0.67	0.53,0.85	0.67	0.53,0.85
Intermediate / poor risk subgroup OS				
cabozantinib	0.8	0.53,1.21	0.81	0.57,1.21
Pembrolizumab+axinib	0.52	0.37,0.74	0.53	0.37,0.74

Shaded cells indicate disagreement between CS and ERG results