Treatments for renal cell carcinoma [ID6186]: A Pathways Pilot Appraisal Assessment Report

Produced by

Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU

Authors

Dawn Lee[1]
Madhusubramanian Muthukumar[1]
Alan Lovell[1]
Caroline Farmer[1]
Darren Burns[2]
Justin Matthews[1]
Helen Coelho[1]
Brian O'Toole[1]
Laura A Trigg[1]
Tristan Snowsill[1]
Maxwell S Barnish[1]
Thalia Nikoglou[1]
Amanda Brand[1]
Zain Ahmad[1]
Ahmed Abdelsabour[1]
Louise Crathorne[1]
Sophie Robinson[1]
Edward CF Wilson[1]
G.J. Melendez-Torres[1]

[1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter, UK

[2] Delta Hat Limited, Long Eaton, Nottingham, UK

Correspondence to Professor Dawn Lee

3.09 South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; D.Lee7@exeter.ac.uk

Date completed 13/04/2023

Dated updated 24/08/2023

Source of funding This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR136008.

Declared competing interests of the authors None, however, Professor Larkin also provided clinical input to NICE during the scoping stages of the appraisal.

Acknowledgments The authors acknowledge the clinical expert input provided by Prof James Larkin (Royal Marsden) in the development of this report and the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe (both PenTAG). Additional support was also provided by Mr Ben Shaw and Dr David Salmon.

Rider on responsibility for document The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows Lee et al. Treatments for renal cell carcinoma [ID6186]: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2023.

Author Contributions:

Dawn Lee	Project lead, lead for production of the decision model, input to design of observational data analysis, conduct of the economic literature searches, writing and editorial input
Madhusubramanian Muthukumar	Development of excel interface for the decision model inputs, screening and data extraction of evidence identified in the economic literature searches and input to production of the decision model, economic writing
Alan Lovell	Production of the literature search strategies, identification of potential observational data sources and contact with relevant data holders, writing the literature search methods
Caroline Farmer	Screening evidence identified in the clinical literature searches, identification of clinical experts for consultation, quality assurance of clinical data extraction, production of write up of clinical evidence and treatment pathway
Darren Burns	Production of the decision model
Justin Matthews	Conduct and writeup of fractional polynomial network meta-analysis. Input to design of observational data analysis

Helen Coelho	Screening evidence identified in the clinical literature searches, clinical data extraction, production of write up of treatment pathway
Brian O'Toole	Screening and extracting evidence identified in the economic literature searches, analysis of publicly available observational data, economic writing
Laura A Trigg	Screening evidence identified in the economic literature searches, data extraction of economic data, economic writing, data cleaning for observational dataset
Tristan Snowsill	Production and QC of the R decision model
Thalia Nikoglou	Proof reading and screening of clinical data, recruitment of clinical experts
Maxwell S Barnish	Clinical data extraction, review of quality assessment of clinical trials, clinical writing
Amanda Brand	Quality assessment of clinical trials
Louise Crathorne	Review of real-world evidence, patient and clinical commentary sections and proof reading
Zain Ahmad	Clinical and economic data extraction, and production of expert elicitation survey
Sophie Robinson	Advising on literature search methods, quality assurance of the literature search strategies
Edward CF Wilson	Production of the decision model
G.J. Melendez-Torres	Guarantor of the report. Conduct of statistical analysis, screening and data extraction for clinical literature searches and production of write up of clinical evidence and treatment pathway

All [redacted] data has been highlighted in blue and underlined, all [redacted] data is highlighted in yellow and underlined

Table of Contents

List of Tables

List of Figures

List of Key Issues

List of Abbreviations

Acronym	Definition
ABPI	Association of the British Pharmaceutical Industry
AE	Adverse event
AG	Assessment group
AIC	Akaike information criterion
ALT	Alanine aminotransferase
aRCC	Advanced RCC
ASCO	American Society of Clinical Oncology
AST	Aspartate aminotransferase
AUC	Area under the curve
BIC	Bayesian information criterion
BICR	Blinded independent central review
BID	Twice daily
BM	Bone metastases
BMJ	British Medical Journal
BMS	Bristol Myers Squibb
BNF	British National Formulary
BRL	Brazilian Real
BSC	Best supportive care
CADTH	Canadian Agency for Drugs and Technologies in Health
ccRCC	Clear cell renal cell carcinoma
CDF	Cancer Drugs Fund
CDSR	Cochrane Database of Systematic Reviews
CEA	Cost-effectiveness analysis
CI	Confidence interval
CMS	Centers for Medicare and Medicaid Services
CNS	Central nervous system
CPI	Consumer Price Index
CPRD	Clinical Practice Research Datalink
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
CT	Computed tomography
DAPS	Direct access pathology services
DBL	Database lock
DCR	Disease control rate
DES	Discrete event simulation
DF	Degrees of freedom
DFS	Disease-free survival
DIC	Deviance information criterion

Acronym	Definition
DICE	Discretely integrated condition event
DM	Distance metastases
DoR	Duration of response
DSU	Decision Support Unit
DVT	Deep vein thrombosis
EAG	External assessment group
ECOG	Eastern Cooperative Oncology Group
EED	Economic Evaluation Database
EHR	Electronic health record
EMA	European Medicines Agency
eMIT	Drugs and pharmaceutical electronic market information tool
EPAR	European public assessment report
ERG	Evidence review group
ESMO	European Society for Medical Oncology
EUDRACT	European Union Drug Regulating Authorities Clinical Trials Database
FACT	Functional Assessment of Cancer Therapy quality of life questionnaire
FAD	Final appraisal determination
FDA	U.S. Food and Drug Administration
FE	Fixed effects
FKSI-DRS	Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index - Disease Related Symptoms
FP	Fractional polynomials
G-BA	Federal Joint Committee (Germany)
GBP	Great British Pounds
GFR	Glomerular filtration rate
GG	Generalised gamma
GP	General practitioner
GU	Genito urinary
HAS	Haute Autorité de Santé (France)
HCRU	Healthcare resource use
HES	Hospital Episode Statistics
HFS	Hand foot syndrome
HR	Hazard ratio
HRG	Health resource group
HRQoL	Health-related quality of life
HSE	Health Survey England
HTA	Health technology assessment
IA	Investigator assessment
ICER	Incremental cost-effectiveness ratio
ICI	Immune checkpoint inhibitor
ICTRP	International Clinical Trials Registry Platform
IFN	Interferon

Acronym	Definition
IL	Interleukin
IMDC	International Metastatic RCC Database Consortium
INAHTA	International Network of Agencies for Health Technology Assessment
INR	International normalized ratio
IO	Immuno-oncology
IPD	Individual patient-level data
IPW	Inverse probability weighting
IQR	Interquartile range
IRIN	irinotecan
IRRC	Independent radiology review committee
IRT	Interactive Response Technology
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITT	Intention-to-treat
IV	Intravenous
IVI	Innovation and Value Initiative
IxRS	Interactive voice or web response systems
KM	Kaplan Meier
KPS	Karnofsky performance status
LDH	Lactate dehydrogenase
LOCF	Last observation carried forward
LOT	Line of treatment
LR	Local recurrence
LVEF	Left ventricular ejection fraction
LYG	Life years gained
MA	Meta-analyses
MCM	Mixture-cure model
MDG	Modified de Gramont
MHRA	Medicines and Healthcare products Regulatory Agency
MMRM	Mixed model repeated measures
MoA	Mechanism of action
MRC	Medical Research Council
mRCC	Metastatic renal cell carcinoma
MSKCC	Memorial Sloan Kettering Cancer Center
MTA	Multiple technology appraisal
NA	Not applicable
NB	Net benefit
NCCN	National Comprehensive Cancer Network
nccRCC	Non-clear cell renal cell carcinoma
NCRAS	National Cancer Registration and Analysis Service
NDRS	National Disease Registration Service
NE	Not evaluable
NEJM	New England Journal of Medicine

Acronym	Definition
NHS	National Health Service
NHSCII	NHS Cost Inflation Index
NHSE	National Health Service, England
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
NMA	Network meta-analysis
NOS	Not otherwise specified
NR	Not reported
NSCLC	Non-small cell lung cancer
OD	Once daily
OLS	Ordinary least squares
ONS	Office for National Statistics
OR	Odds ratio
ORR	Overall response rate
OS	Overall survival
OWSA	One way sensitivity analysis
PartSA	Partitioned survival analysis
PAS	Patient access scheme
PD	Progressive disease
PF	Progression free
PFS	Progression-free survival
PH	Proportional hazards
PICOS	Population, Intervention, Comparison, Outcomes and Study
PO	Per os (orally)
PPS	Post progression survival
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient reported outcome
PS	Performance status
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life-year
QC	Quality control
QD	Every day
RCC	Renal cell carcinoma
RCT	Randomised controlled trial
RDI	Relative dosing intensity
RE	Random effects
RECIST	Response Evaluation Criteria in Solid Tumours
REMARCC	Registry for Metastatic Renal Cell Carcinoma
RWD	Real-world data
RWE	Real-world evidence

Acronym	Definition
SABR	Stereotactic ablative radiotherapy
SACT	Systemic Anti-Cancer Therapy
SCLC	Small-cell lung cancer
SD	Standard deviation
SE	Standard error
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic literature review
SmPC	Summary of Product Characteristics
STA	Single technology appraisal
TA	Technology appraisal
TE	Treatment effect
TKI	Tyrosine kinase inhibitor
TSD	Technical support document
TTD	Time to treatment discontinuation
TTNT	Time to next treatment
ToT	Time on treatment
TTP	Time to progression
TTR	Time to response
TWSA	Two-way sensitivity analysis
UK	United Kingdom
ULN	Upper limit of normal
US	United States
VAS	Visual Analogue Scale
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
WHO	World Health Organization
ZIN	Zorginstituut Nederland National Health Care Institute

Key issues summary

The decision problem: summary of the EAG's key issues

Key Issue 1: Optimal sequencing of treatments, including after novel first-line treatments

Report sections	
Description of issue and why the EAG has identified it as important	Clinical advice to the EAG and consideration of relevant evidence highlights that optimal treatment sequencing following novel treatments at first line (i.e. IO/IO or IO/TKI combinations) remains an area of uncertainty. In addition, evidence for optimal treatment choice and sequencing in favourable risk patients at first-line remains an area of clinical debate.
What alternative approach has the EAG suggested?	The EAG has received clinical advice as to most likely treatment sequences. However, additional clinical evidence is needed to ascertain which treatments are most likely to be received, and most effective, as novel treatments continue to emerge in first line; as well as optimal treatment choice for favourable risk patients.
What is the expected effect on the cost-effectiveness estimates?	Current estimates of cost effectiveness, particularly in second line and for favourable risk patients, may evolve as this evidence develops. Optimal treatment sequencing may also impact overall estimates of OS in first line, but the direction of impact on cost-effectiveness estimates is unclear.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 2: Company's definition of relevant comparators

Report sections	
Description of issue and why the EAG has identified it as important	The company argued that, at first line, avelumab plus axitinib is a relevant comparator, and excluded tivozanib. The EAG disagrees with this position as avelumab plus axitinib is not considered to be routinely commissioned while it is accessed through the Cancer Drugs Fund; further, tivozanib is a relevant treatment at first line.
What alternative approach has the EAG suggested?	The EAG has included avelumab plus axitinib in clinical effectiveness analyses for completeness in line with the scope of the pathways decision problem (rather than the decision problem specific to cabozantinib plus nivolumab), but has not included this treatment in economic analyses for cabozantinib + nivolumab in keeping with NICE guidance. The EAG has also included tivozanib where possible in first-line analyses acknowledging limitations in the ability to conduct indirect treatment comparisons.
What is the expected effect on the cost-effectiveness estimates?	The EAG's cost-effectiveness estimates will more closely reflect NICE guidance.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 3: Company's definition of relevant outcomes

Report sections	
Description of issue and why the EAG has identified it as important	The company argued in its original submission that time to next treatment was not a relevant outcome. When these data were provided, the definition used was non-standard, precluding meaningful comparisons to other studies.
What alternative approach has the EAG suggested?	The EAG has suggested defining time to next treatment in a way similar to other studies; i.e. considering the time from initiation of first-line treatment to the first of uptake of a second systemic treatment where this has been recorded, death or loss to follow-up. These data are not yet available.
What is the expected effect on the cost-effectiveness estimates?	The EAG's economic modelling will be able to draw on data for this outcome to produce more consistent and high-fidelity cost-effectiveness estimates.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 4: Company's definition of relevant subgroups

Report sections	
Description of issue and why the EAG has identified it as important	The company argued in its original submission that cabozantinib plus nivolumab should be assessed in the all-risk group. The EAG notes that risk group is known to be an important prognostic factor, an important effect modifier across a range of RCC treatments, and a key factor in previous NICE appraisals, as well as a salient factor in clinical decision-making. As a result, subgroup-specific evidence is highly probative. Moreover, in subgroup-specific network meta-analyses, the EAG found that patterns of effect were different by risk group.
What alternative approach has the EAG suggested?	The EAG has considered cost-effectiveness both in an all-risk population as well as in intermediate/poor risk populations and favourable risk populations separately, reflecting practice in prior appraisals for RCC.
What is the expected effect on the cost-effectiveness estimates?	The EAG expects that cost-effectiveness estimates will more closely reflect clinical realities and the existing treatment pathway, supporting more robust decision-making.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

The clinical effectiveness evidence: summary of the EAG's key issues

Key Issue 5: CheckMate 9ER: Consistency of reporting

Report sections	
Description of issue and why the EAG has identified it as important	The company submitted an interim report of clinical effectiveness, with a subsequent update provided due to data quality issues. However, the EAG did not find that the explanation of changes provided was sufficiently comprehensive to provide confidence in the data quality. For example, data relating to adverse events had minor changes that were not explicitly described as updated.
What alternative approach has the EAG suggested?	It was not possible for the EAG to resolve this issue within its appraisal using the available data. A clear explanation of all changes made

Report sections	
	between data cuts provided would increase confidence in the analyses provided.
What is the expected effect on the cost-effectiveness estimates?	It is unclear if an explanation would impact data inputs to the EAG's economic model; however, confidence in data quality is essential to minimise decision risk.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 6: CheckMate 9ER: Generalisability of the trial to UK practice

Report sections	
Description of issue and why the EAG has identified it as important	The EAG's inspection of the company's trial data found that the trial enrolled a relatively small number of UK patients, and that the rate of patients continuing to receive treatment post-progression was both higher than expected and not in keeping with clinical treatment patterns in the UK. In addition, patients with intermediate and poor risk receiving sunitinib had higher restricted mean survival times for both OS and PFS in the CheckMate 9ER trial than the comparable real world evidence source preferred by the EAG, with a similar trend seen for OS in the favourable risk group as well. Patients receiving sunitinib also had comparatively lower use of nivolumab as a subsequent treatment than expected.
What alternative approach has the EAG suggested?	It was not possible for the EAG to resolve this issue within its appraisal using the available data. A clearer justification of why post-progression treatment rates were higher than expected would contextualise concerns about generalizability. Analyses accounting for post-progression treatment would be valuable to better understand the impact of post-progression treatment rates, and mix of post-progression treatments.
What is the expected effect on the cost-effectiveness estimates?	Clearer understanding of time on treatment post-progression would impact treatment costs estimated in an economic model. The direction of this impact is unclear pending an explanation from the company.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 7: CheckMate 9ER: Effect modification by risk group

Report sections	
Description of issue and why the EAG has identified it as important	The EAG's inspection of the company's trial data found that there was some evidence of effect modification by risk group for OS and PFS; for example, the hazard ratio for OS comparing cabozantinib plus nivolumab against sunitinib in favourable-risk patients (HR=1.07) is more than twice as high as for patients with poor risk (HR=0.46), with a similar trend in evidence for PFS (HR=0.72 vs HR=0.37). This is important because it reinforces the value of risk group as a key consideration in this appraisal and its salience in clinical and cost-effectiveness decision-making.
What alternative approach has the EAG suggested?	The EAG reiterates that cost-effectiveness modelling should also consider risk group as a key factor, including production of cost-effectiveness estimates by risk group.
What is the expected effect on the cost-	Estimates for the cost-effectiveness of cabozantinib plus nivolumab are likely to be very different by risk group.

Report sections	
effectiveness estimates?	

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 8: Evidence base: quality and sufficiency of included randomised trials

Report sections	
Description of issue and why the EAG has identified it as important	The EAG's appraisal of the randomized trials included in its syntheses identified significant limitations in the quality of included trials, including CheckMate 9ER; of the 17 prioritised trials, nine were appraised as being at high risk of bias and eight were appraised as being at an unclear risk of bias. The majority of comparisons in first-line and second-line networks were informed by only one trial, meaning that many comparisons between novel treatments were based on indirect evidence only, and inconsistency in networks could not be assessed. Moreover, risk group-specific analyses drew on comparatively sparse data, which were often unevenly presented; in particular, pembrolizumab plus lenvatinib could not be included in risk group-specific fractional polynomial NMAs for PFS due to redacting of data in TA858.
What alternative approach has the EAG suggested?	The EAG has used parallel analysis methods for survival outcomes, including fractional polynomial NMA and proportional hazards NMA, to test the robustness of analyses to different assumptions where possible. However, only proportional hazards NMAs are available for survival outcomes in the favourable risk group patients in first line. However, this does not address the challenges relating to risk of bias.
What is the expected effect on the cost-effectiveness estimates?	Estimates for the cost-effectiveness of cabozantinib plus nivolumab are increased in their statistical uncertainty due to limitations and sparseness in the underpinning evidence base; in addition, it is impossible to quantify the impact of trial-level bias on cost-effectiveness estimates.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 9: Evidence base: distribution of effect modifiers across evidence networks

Report sections	
Description of issue and why the EAG has identified it as important	While the EAG did not regard that distribution of effect modifiers across the network precluded the feasibility of NMAs, it remains that differences between trials in risk group distribution, histological features, proportion with prior nephrectomy, proportion with sarcomatoid features and, to a possibly lesser degree, age could not be meaningfully addressed in NMAs. This was both because of the sparseness of networks and because of poor reporting of several of these characteristics (particularly proportion with sarcomatoid features). More generally, observational evidence suggests that over time and in the last 15 years, patients have experienced better outcomes regardless of treatment. Trials included draw from a wide range of timeframes and follow-up lengths, adding another challenge to interpretation.
What alternative approach has the EAG suggested?	The EAG used a random effects term when appropriate in its fractional polynomial NMAs, which accounted for some heterogeneity in baseline risk. However, a network meta-regression with a less sparse evidence network would have provided greater confidence in findings.

Report sections	
What is the expected effect on the cost-effectiveness estimates?	The direction of travel of cost-effectiveness estimates as a result of this uncertainty is difficult to quantify, as it in part depends on the age of the trial and trial-specific distribution of effect modifiers. However, given lower numbers of poor risk patients in trials linking tivozanib in first-line networks, estimates may be biased in favour of tivozanib.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 10: Evidence base: non-proportional hazards and evolution over time in survival outcomes

Report sections	
Description of issue and why the EAG has identified it as important	Many of the prioritised trials exhibited violations of the proportional hazards assumptions, based either on statistical tests or on visual inspection. In addition, time-to-event data were drawn from the last available data cut given difficulties in identifying 'most similar' time points for analysis and to avoid discarding collected data. However, differential trial maturity is a challenge for interpretation given evidence of 'slippage' in HRs towards the null, particularly for IO/TKI combinations, over sequential follow-ups.
What alternative approach has the EAG suggested?	As above, the EAG has used parallel analysis methods for survival outcomes, including fractional polynomial NMA and proportional hazards NMA, to test the robustness of analyses to different assumptions. However, challenges in estimating hazard functions generated some inconsistencies between both analysis strategies, particularly for pembrolizumab plus lenvatinib in first-line, and generated estimates for second-line fractional polynomial NMAs that were inconsistent between outcomes. It is likely that the EAG's analyses should be revisited when all trials have reached maturity.
What is the expected effect on the cost-effectiveness estimates?	Based on evidence of slippage, it is likely that cost-effectiveness estimates for novel treatments drawing on comparatively less mature trials may be unduly optimistic.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

Key Issue 11: Evidence base: unanswered questions relating to applicability across histologies and in a context of adjuvant treatment

Report sections	
Description of issue and why the EAG has identified it as important	Included trials primarily restricted inclusion to patients with clear cell RCC, creating questions about the applicability of analyses to other RCC histologies. In addition, adjuvant pembrolizumab is now available in routine practice, but was not available as part of routine practice when any of the included trials were conducted. Clinical advice to the EAG is that adjuvant pembrolizumab may reduce the subsequent effectiveness of IO treatments and improve prognosis for other types of treatment as patients will be scanned more regularly, leading to earlier detection and treatment of progression.
What alternative approach has the EAG suggested?	The EAG could not address these issues in this appraisal due to sparsity of evidence. However, a number of trials are emerging in different RCC histologies which will provide additional evidence in this area.

Report sections	
What is the expected effect on the cost-effectiveness estimates?	As adjuvant pembrolizumab increases in use, it is likely that effect estimates from IO treatments will vary in practice from those observed in key trials. These may eventually attenuate the cost-effectiveness of IO-based treatments, particularly in first line.

Abbreviations: EAG, External Assessment Group; HR, hazard ratio; IO, immuno-oncology treatment; NMA, network meta-analysis; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor

1. OBJECTIVES OF THE PILOT PROCESS AND THIS ASSESSMENT

The NICE Pathways pilot process aims to enhance the efficiency of assessing treatments and inform access decisions by developing a comprehensive and adaptable core model for specific disease areas.

NICE selected RCC as the first pilot topic due to the expected pipeline of treatments, indicating a dynamic and evolving landscape in RCC therapies. RCC is a disease area characterised by multi-comparator decision spaces, meaning there are several treatment options available at different stages of the disease pathway. Treatment decisions in RCC are influenced by factors such as the patient's exposure to prior therapies, disease progression, and individual patient characteristics. The NICE Pathways pilot process for RCC seeks to test an evaluation framework that can effectively assess and compare various treatment options within the RCC pathway. By considering the evolving landscape of RCC therapies, the process aims to inform access decisions, optimise treatment pathways, and ultimately benefit patients with RCC.

As part of this pilot NICE requested the development of an EAG model which incorporates multiple decision nodes to assess multiple technologies in a disease pathway and inform robust access decisions. NICE has published a process statement outlining the summary of this pilot and the intended process to achieve its aims.[1] Within this pilot the aim was to develop a high-quality open-source disease model, available to all relevant stakeholders without restriction, which can be reused and built upon in future appraisals whilst maintaining confidentiality of proprietary data.

An attractive model for this type of approach is the Innovation and Value Initiative's Open-Source Value Project (IVI; Jansen et al. 2019[2]). Since the project began in 2018, IVI has developed three disease models – one in rheumatoid arthritis, one in non-small-cell lung cancer and one in major depressive disorder – that are made freely available to all users, with full open-source code posted in a public repository (GitHub).[3] As part of its development process, IVI holds regular public consultation seeking feedback on the structure and parameterisation of its analyses, and exposing its implementation to unrestricted scrutiny.

Given the scope and steps of the process the consultation stage is different to the IVI models. In particular, a user-interface will not be provided prior to the Appraisal Committee meeting and is scheduled instead for a later phase of work (see Section 4.3.1.9). However, the code will be posted in a public repository enabling full public scrutiny and as discussed additional functionality will be incorporated during Phase 2 of the pilot.

2. DECISION PROBLEM, DESCRIPTION OF THE TECHNOLOGIES AND CLINICAL CARE PATHWAY

2.1. Description of the health condition

RCC is a cancer that usually originates in the lining of the tubules of the kidney (the smallest tubes inside the nephrons) that help filter the blood and make urine. RCC is the most common type of kidney cancer, accounting for more than 80% of cases.[4] Clear cell RCC is the most common subtype, quoted as accounting for approximately 75% of cases.[4] Non-clear cell subtypes have varying frequencies, with papillary RCC comprising around 10-15% of cases, chromophobe RCC around 5%, and other subtypes representing approximately 1% each.[4] Non-classifiable RCC is a rare category and histological assessment can be challenging, especially when a nephrectomy was not performed initially or limited tissue samples are available.

Diagnosis is usually incidental, and when people present with symptoms the disease is usually advanced; the most common symptoms being upper abdomen or back pain, a palpable lump or mass in the kidney area and gross haematuria.[5,6] In metastatic disease, symptoms associated with the metastatic tumours may be present, including airway obstruction, bleeding, and dyspnoea for lung metastases, pain and fractures for bone metastases, jaundice and swelling for liver metastases, and swelling of lymph nodes for lymphatic metastases.[7-9]

RCC is typically staged from Stage 1 to Stage 4 according to how far the cancer may have spread; Stage 3 indicates that the cancer has advanced locally (within regional lymph nodes) and Stage 4 indicates that metastases beyond the regional lymph nodes are present. Treatment depends on the location and stage of the cancer.[10]

The scope for this appraisal is people with advanced RCC (aRCC) or metastatic RCC (mRCC). Although systemic treatments are mostly suitable for those with metastatic disease (Stage 4), they may be offered to people with locally advanced (Stage 3) disease where this is unresectable. Due to this, people with Stage 4 RCC or Stage 3 unresectable RCC have been included in this appraisal.

2.2. Epidemiology

Kidney cancer is the eighth most common cancer in the UK, accounting for 4% of all new cancer cases (2019).[11,12] Kidney cancer is more common in men than in women: in the UK, between 2016 and 2018, there were 1.7 times more new cases in men than in women. A quarter of cases were diagnosed in people aged 60 to 69 years, with nearly half of cases (49.7%) diagnosed in people aged ≥70 years.[11] It is also more common in people of white ethnicity.[11] Links to certain lifestyle factors such as obesity, hypertension and smoking are well-established.[13]

In 2018, 9,438 new kidney cancer cases were diagnosed in England.[14] Of those, 40.2% had Stage 1 disease, 7.6% had Stage 2 disease, 15.5% had Stage 3 disease and 20.5% had Stage 4 disease.[10] The five-year survival has been reported as 86.8%, 76.6%, 74.2% and 12.4% for Stage 1, 2, 3, and Stage 4 disease, respectively.[15] These survival rates are likely to underestimate survival for patients starting treatment now as they do not include the impact of immuno-oncology combinations that have more recently entered clinical practice.

RCC is the most common type of kidney cancer, responsible for more than 80% of all cases diagnosed in the UK.[16,17] In aRCC (Stage 3), the tumour located in the kidney may be any size if it has spread to regional lymph nodes or may have grown into major veins or perinephric tissue but has not spread to other parts of the body.[18] In mRCC (Stage 4), the tumour may have spread to areas beyond Gerota's fascia, extending into the adrenal gland on the same side of the body as the tumour and possibly to lymph nodes, but not to other parts of the body, or has spread to any other organ. Metastases in RCC most commonly occur in the lung, bone, lymph node, and liver, leading to significant morbidity as well as poor prognosis.[9]

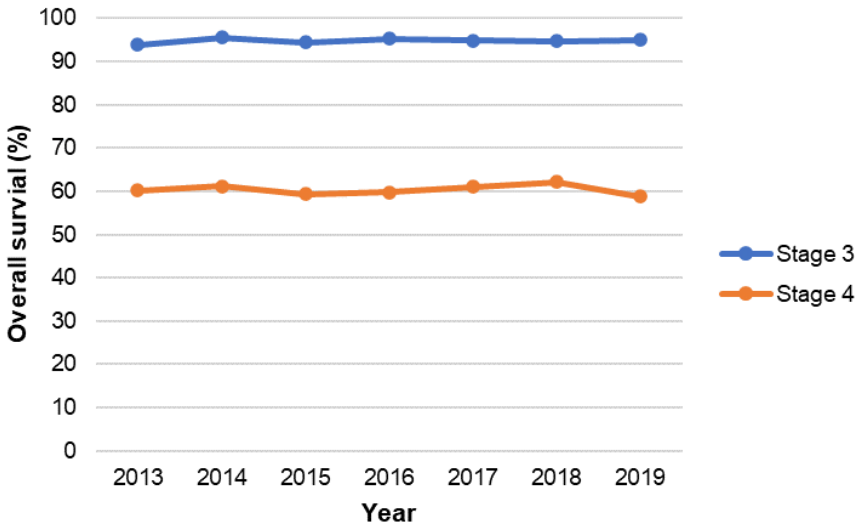
OS data for RCC were available from the Get Data Out (GDO) 'Kidney' dataset, published by the NCRAS. Yearly data (from 2013 to 2019) were recorded for Stage 1, 2, 3 and 4 clear cell RCC patients (n=252), and for RCC not otherwise specified (NOS) (n=364) (patients may be NOS either because a distinct morphology cannot be seen under the microscope [histologically confirmed], or because the tumour has been clinically diagnosed and no tissue sample has been taken [not histologically confirmed]). Survival rates were reported as Kaplan-Meier (KM) estimates at Month 3, 6, 9, 12, 24, 36, 48, 60, 72 and 84. The most complete data were for 12 months i.e. 12-month data were reported for all years. The data

indicate that patients with Stage 3 clear cell RCC have better 12-month prognosis/ highest survival rates (ranging from 93.9% to 95%) than those with Stage 3 or 4 RCC with any other histological profile. The majority of these patients will not be eligible for surgery and therefore not in scope of this appraisal.

Stage 4 clear cell RCC, is the histology in which the majority of clinical trials have been conducted. Cancer Research UK data indicate that this makes up approximately 75% of all RCC cases, whereas NCRAS data indicate that this makes up 77 % of Stage 3 and 44% of Stage 4 RCC case.[16,19] Clinical expert advice has indicated that 44% clear cell at Stage 4 is lower than typically seen in clinical practice suggesting that some patients may not have undergone a biopsy. For Stage 4 clear cell patients, 12-month survival ranged from 58.5% to 62.2% (Figure 1 and Figure 2). The most severe histological subtype with the lowest 12-month overall survival estimates were patients with Stage 4 renal cell carcinoma NOS (not histologically confirmed), ranging from 13.1% to 18.4%.

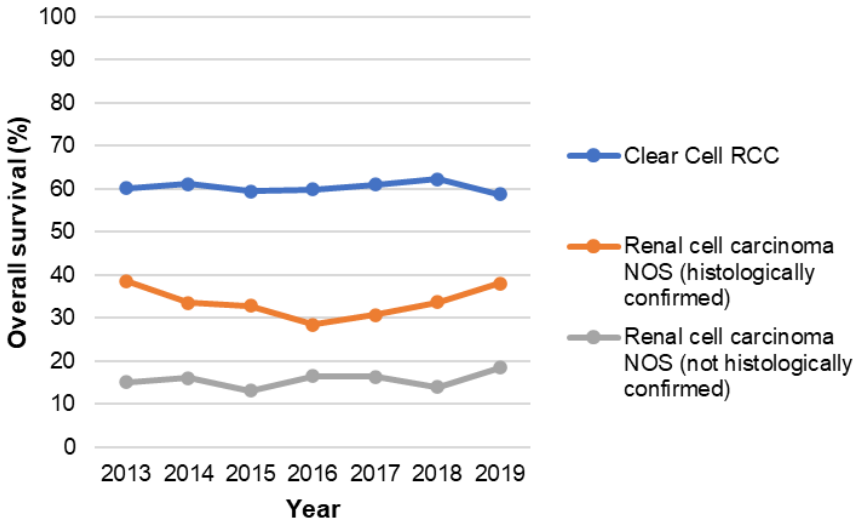
The data suggest that there has been a sustained improvement in OS from 2016 to 2019 for patients with Stage 4 RCC NOS (histologically confirmed), with OS increasing from 28.5% to 38%. Although the cause for improved survival rates is not clear, it may be due to patient enrolment in clinical trials focusing on non-clear cell histologies.

Figure 1: 12-month overall survival for Stage 3 and 4 clear cell RCC (2013-2019)



Abbreviations: RCC, renal cell carcinoma

Figure 2: 12-month overall survival for Stage 4 cancer, all histologies (2013-2019)

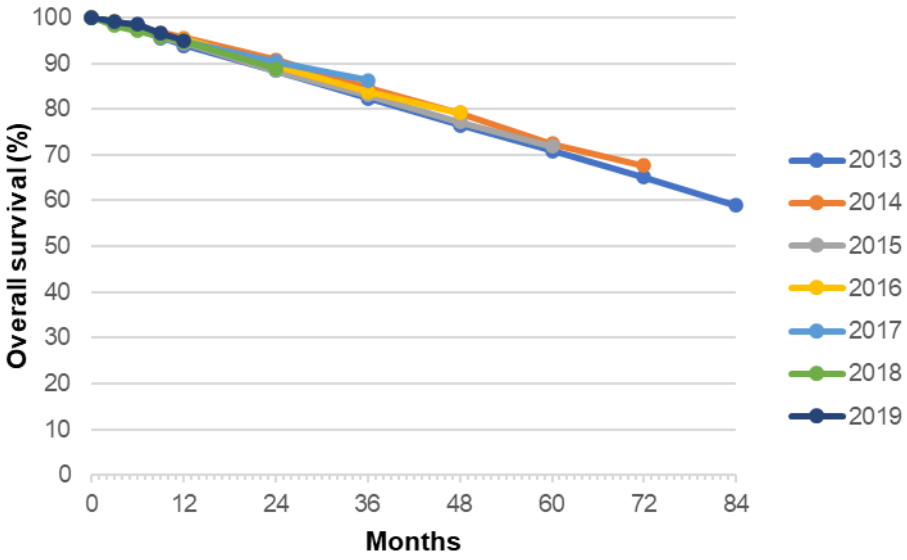


Abbreviations: RCC, renal cell carcinoma

Five-year (60 month) survival rates were recorded for years 2013, 2014 and 2015. For completeness and for validation purposes these are outlined below. OS at 60 months confirm that patients with Stage 3 clear cell RCC have the best 12-month prognosis/ highest survival rates (ranging from 70.8% to 72.4%). For Stage 4 clear cell RCC, 60-month survival ranged from 19.1% to 20.1%. Patients with Stage 4 RCC NOS (not histologically confirmed) have the poorest 12-month prognosis/lowest survival rates (ranging from 2.1% to 2.7%).

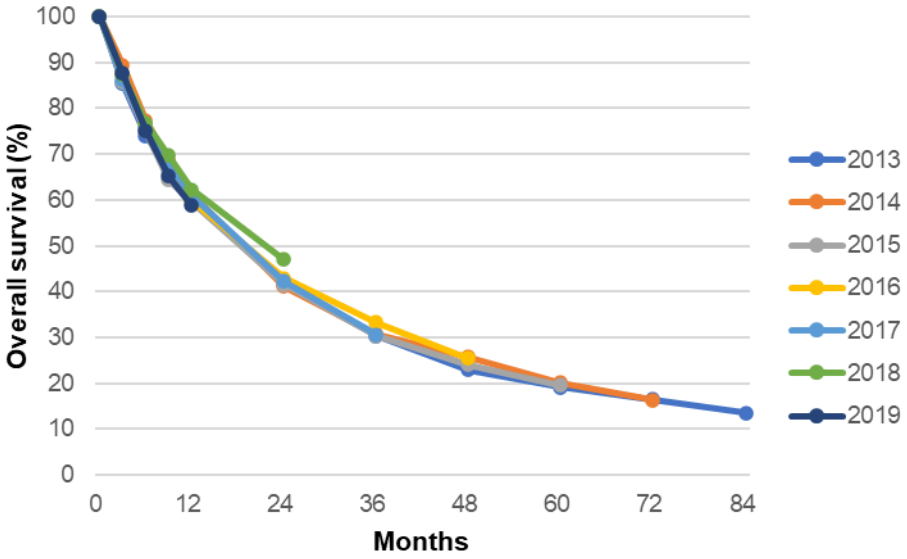
Figure 3 and Figure 4 show that the prognosis for clear cell RCC remained relatively consistent between 2013 and 2019, however, as noted earlier these survival rates are likely to underestimate survival for patients starting treatment now as they do not include the impact of immuno-oncology combinations that have more recently entered clinical practice for which any improvements are most likely to be seen in longer-term data.

Figure 3: Overall survival for patients with Stage 3 clear cell RCC (all years)



Abbreviations: RCC, renal cell carcinoma

Figure 4: Overall survival for patients with Stage 4 clear cell RCC (all years)



Abbreviations: RCC, renal cell carcinoma

2.3. Prognostic factors

Prognostic factors play a key role in aRCC by providing valuable insights into disease prognosis and guiding treatment decisions. Several important prognostic factors have been identified in aRCC.

Risk scores, such as the International Metastatic RCC Database Consortium (IMDC) and Memorial Sloan Kettering Cancer Center (MSKCC) scores, are widely used tools that incorporate various factors including performance status, time from diagnosis to systemic therapy initiation, haemoglobin levels, calcium levels, and lactate dehydrogenase (LDH) levels. These scores help classify patients into favourable, intermediate, and poor-risk groups, providing valuable information about disease aggressiveness and treatment response (see Section 2.3.1).

Histology is another key prognostic factor, with clear cell RCC being the most common subtype and generally associated with a poorer prognosis compared to other subtypes.[20] The presence of metastasis is a well-established prognostic factor in aRCC, indicating the extent and aggressiveness of the disease.[20] Differentiating between visceral metastases and bone metastases is also important, as patients with bone metastases often exhibit a less favourable outcome and suboptimal response to certain treatments, such as tyrosine kinase inhibitors (TKIs).[20]

Nephrectomy is an additional prognostic factor in aRCC. In select patients, nephrectomy has shown benefits, especially in favourable-risk disease, with improved survival compared to those who do not undergo the procedure. In cases where nephrectomy is performed, it typically indicates that the primary tumour was localised and surgically resectable. This suggests that the disease had not spread extensively beyond the kidney at the time of diagnosis. Consequently, patients who undergo nephrectomy in these circumstances tend to have a more favourable prognosis compared to those with primary metastatic disease.[21] On the other hand, if a patient presents with primary metastatic disease, nephrectomy may not be pursued as the cancer has already spread beyond the kidney to other distant sites. The presence of metastasis often indicates a more advanced stage of the disease, and the prognosis for such patients tends to be poorer.[21]

Timely initiation of systemic therapy is also a significant prognostic factor, as delayed treatment may adversely affect outcomes. Patients who received treatment within 100 days of diagnosis had a lower OS from the start of systemic

treatment compared to those who initiated treatment 600 days or more after diagnosis.[22] Early intervention with targeted therapies or immunotherapies has been associated with better response rates and prolonged survival.

Sarcomatoid features within the tumour represent another important prognostic factor in aRCC.[20,23] Sarcomatoid RCC, characterised by spindle or giant cells resembling a sarcoma, is associated with a poorer prognosis. This variant often exhibits larger tumour size, extensive disease, and a higher likelihood of metastasis. Additionally, sarcomatoid differentiation can lead to resistance against systemic therapies, limiting treatment options and reducing overall survival rates.

Other prognostic factors in aRCC include age, tumour stage, PS,[24,25] and laboratory parameters such as haemoglobin levels, LDH levels, and calcium levels.[26] These parameters provide additional information about disease aggressiveness and can aid in treatment decision-making.

By considering these prognostic factors, clinicians can better evaluate disease prognosis, select appropriate treatment strategies, and optimise outcomes for patients with aRCC.

2.3.1. Risk status

According to expert advice received, risk status for people with aRCC who have not received systemic therapy is classified using the International Metastatic RCC Database Consortium (IMDC) risk score.[27,28] This scoring system was derived from a population of patients with metastatic RCC treated with VEGF-targeted therapy and predicts survival based on time from diagnosis, Karnofsky performance status, and laboratory measures of haemoglobin, corrected calcium and neutrophils. Within the current treatment pathway for RCC, some treatments are only recommended for people with IMDC poor or intermediate risk status (Section 2.4). Although the relevance of IMDC prognostic criteria to frontline combination immunotherapy is still being established, these criteria are commonly used to risk-stratify patients in clinical trials and guide treatment decisions in practice.[29]

Historically, risk status was classified using another risk stratification score: the Memorial Sloan Kettering Cancer Center (MSKCC) risk score,[24,30] which was later extended to create the IMDC system to enhance its predictive accuracy. The IMDC risk score includes additional factors like absolute neutrophil count and

platelet count, which are not considered in the MSKCC model. In UK clinical practice, the IMDC risk score is preferred over the MSKCC risk score because lactate dehydrogenase concentration, which is not routinely tested in the UK, is included in the MSKCC risk score. Studies have shown a high concordance rate (83%) between the two risk scores, with disagreements primarily observed in classifying patients as intermediate or poor risk.[28,31] However, for the purpose of this appraisal, these differences are likely to have limited impact as these groups are generally combined within NICE recommendations.

In UK practice, the majority of patients with RCC are classified as intermediate or poor risk. Recent real-world data indicate rates from 59% to 89% with intermediate or poor risk status on the MSKCC risk score[32,33] and from 69% to 86% using the IMDC risk score.[34,35] Clinical expert advice indicated that approximately 70-75% of RCC patients in the UK have intermediate or poor IMDC risk status, and 25-30% are categorised as favourable risk.

Validation studies have demonstrated that different risk statuses are associated with varying median OS rates. An international study validating the IMDC score reported by Gore et al. in 2015 reported a median OS of 45.5 months for favourable risk, 18.9 months for intermediate risk and 6.2 months for poor risk using data from 4,065 participants between 2004 and 2010.[33] Another study by Yip et al. in 2017,[36] investigating real-world outcomes of 255 individuals treated with immuno-oncology agents, found that while survival data were immature for evaluating 1st line treatment, IMDC risk status was predictive at the 2nd line, with median OS rates not reached, 26.7 months, and 12.1 months ($p<0.0001$) in each of the three risk groups.

Clinical advice suggests that IMDC risk status may be particularly relevant in predicting outcomes for patients receiving treatment with TKIs, as the original risk score was developed using this patient population. Therefore, it is plausible that patients with favourable risk disease may have a higher likelihood of responding to TKI treatment compared to other options.

Risk status is not re-assessed at 2nd or later lines of treatment, and thus the impact of risk assessment on treatment decision-making tends to decrease in subsequent lines. However, it can still be useful for discussing prognosis with patients. In some cases, risk status may be assessed once at the initial diagnosis of aRCC, and that status is carried forward for the patient's subsequent treatment courses. However,

individual patient characteristics and response to treatment may evolve over time, so clinical judgment would be exercised in interpreting and applying risk assessment in later lines of therapy.

2.4. Treatment pathway

The treatment pathway for RCC can be divided into interconnected decision points based on the disease staging system and line of therapy (see Figure 5 and Figure 6). The treatment pathway is based upon people with clear cell histology (as are the majority of trials; Section 3). In practice, the same treatment algorithm is applied to the majority of people with non-clear cell histologies including papillary RCC, chromophobe RCC, collecting duct RCC (Bellini collecting duct RCC), medullary RCC - mucinous tubular and spindle cell RCC, multilocular cystic RCC, XP11 translocation RCC and unclassified RCC.37 Information on the specific histologies where treatments are commissioned in the same manner as clear-cell has been requested from NHSE and will be incorporated into the project findings when received.

2.4.1. Treatment for early stage to locally advanced RCC

Surgery (partial or radical nephrectomy) is usually possible, and is the preferred treatment, for people with early stage to locally advanced RCC and is usually curative.[38] After tumour resection, the cancer can be graded. Risk of recurrence is greater in higher-grade cancers.[39] After surgery, micro-metastases and individual tumour cells may still be present or may reoccur. They can potentially develop into larger tumours and spread to distant sites around the body.[39] This results in advanced, unresectable tumours.[39] The aim of adjuvant treatment is to prevent recurrence and potential progression to advanced (unresectable or metastatic) disease.[40] Approximately 20–40% of people who have received surgery subsequently develop metastatic RCC.[41]

One major change is the introduction of adjuvant treatment. NICE recommended pembrolizumab as an option for the adjuvant treatment of RCC at increased risk of recurrence after nephrectomy, with or without metastatic lesion resection in October 2022.[39] Receipt of pembrolizumab in the adjuvant setting may restrict later treatment options. The reason for this being that the NHS does not fund treatment with subsequent immuno-oncology treatments for people who have received treatment with a programmed cell death protein (PD-1) / programmed death-ligand 1 (PD-L1) inhibitor in the adjuvant setting in the previous 12 months. Based upon expert input patients who are treated in the adjuvant

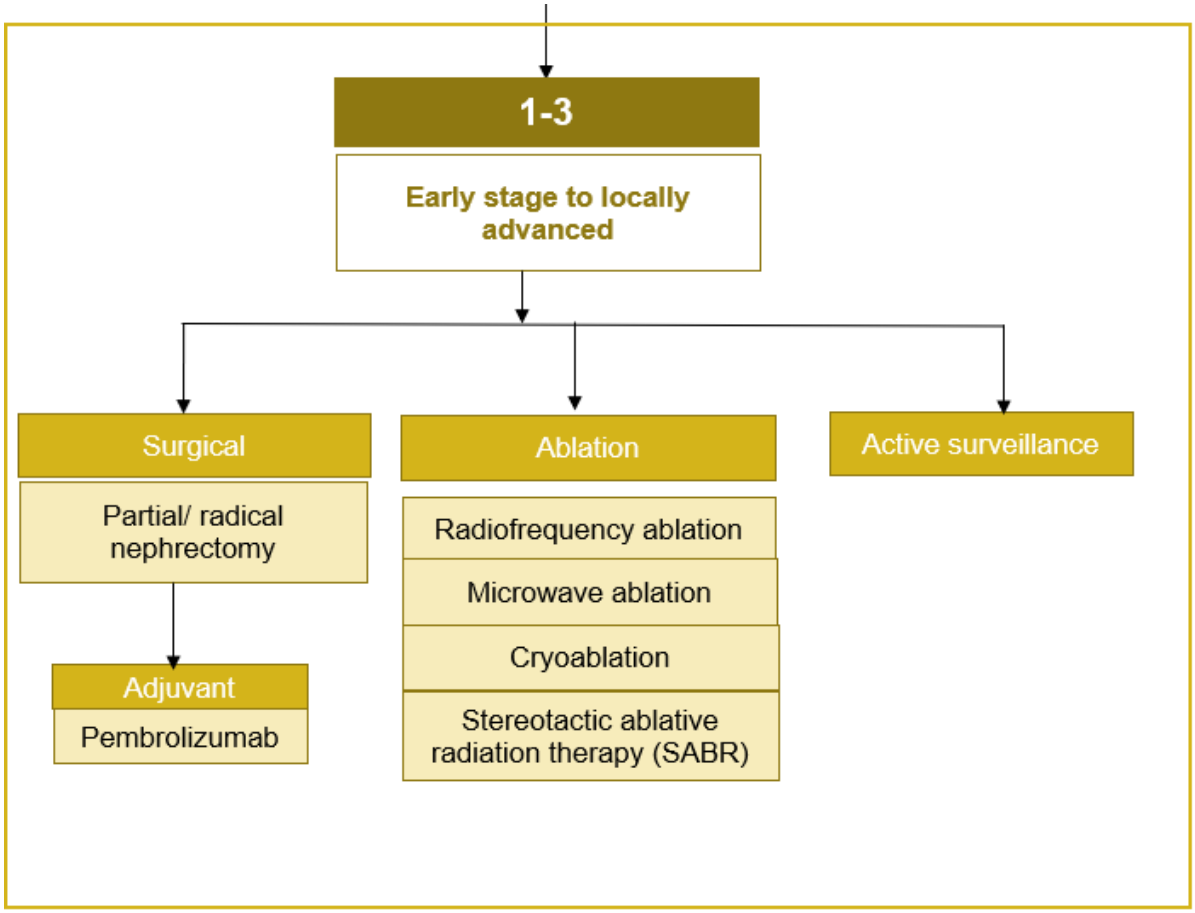
setting are likely to be assessed as favourable risk on IMDC criteria if they relapse as they are scanned frequently which means that relapses are usually detected early.

Clinical feedback to the EAG indicated that the use of adjuvant therapy is a matter of debate among clinicians. While the pembrolizumab trial in the adjuvant setting has reported positive data, trials of other PD-1 inhibitors have reported mixed results. One clinician noted that many clinicians are currently hesitant to use adjuvant treatment due to concerns about toxicity, and the lack of clear selection criteria for identifying patients who would truly benefit from it. In addition, the impact of widespread adjuvant treatment and its effect on relapse rates can significantly influence the validity of existing data. It is still considered too early to determine the uptake of adjuvant pembrolizumab and its impact on the treatment landscape. Currently the proportion of participants receiving adjuvant therapy is low. At the time pembrolizumab was appraised, uptake was expected to start at 20% of the eligible population rising to 65% in five years.[39] Based upon estimates of the eligible population size the maximum uptake is expected to be 18% of the total population. One clinician noted that it will be important to wait for a period of three to six months to assess the real-world utilisation and outcomes of adjuvant pembrolizumab. Understanding the optimal duration and potential long-term effects of adjuvant therapy is crucial for interpreting effectiveness data accurately.

Local ablation is an alternative 1st line approach of particular use in people whose renal function needs to be preserved.[42] The most commonly utilised of these techniques are radiofrequency ablation and cryoablation.[42]

Active surveillance may also be appropriate for early stage RCC, particularly where the mass is small and/or in those who are elderly or frail.[42]

Figure 5: Treatment pathway for early stage to locally advanced RCC



2.4.2. Treatment for advanced RCC

As aRCC is currently incurable, the goal of treatment is to prevent disease progression, maintain health-related quality of life (HRQoL), provide relief from cancer symptoms and extend life.

Treatment guidelines have been developed by the European Society for Medical Oncology (ESMO)[43] and the British Medical Journal (BMJ) RCC best practice guidelines (July 2022).[42] Both guidelines highlight the importance of considering patient factors such as comorbidities, treatment toxicity, and patient preferences when selecting the appropriate treatment regimen. Treatment decisions should be made in consultation with healthcare professionals, taking into account individual patient characteristics and available clinical evidence. While there are no separate NICE guidelines dedicated solely to the management of RCC currently, the NICE recommendations from various technology appraisals (TAs) do guide the treatment of RCC in the UK. Treatments recommended by NICE are summarised in Table 1 and Table 2.

Patients who receive a diagnosis of RCC are afforded a variety of treatment options ranging from active surveillance for those with low volume, indolent disease to cytoreductive nephrectomy for those who for those with favourable outcomes, to treatment with an immune checkpoint inhibitor or tyrosine kinase inhibitor (TKI).

2.4.2.1. Active surveillance or surgery

Treatment options for patients with mRCC include active surveillance and cytoreduction for patients with favourable-risk disease. A subset of patients with mRCC have indolent disease and limited metastatic burden. Initiation of systemic treatment can be postponed in this group of patients to avoid the treatment-related toxicities. In these individuals the ESMO and American Society of Clinical Oncology (ASCO) clinical practice guidelines suggest that active surveillance may be an appropriate option.[43,44] This approach involves closely monitoring the patient's condition without immediate treatment intervention. Active surveillance allows for regular assessments of disease progression and can help avoid unnecessary treatment in patients who may have slower-growing tumours or who may benefit from delayed intervention.

Surgery is only recommended in people where there is a metastasis in a single regional lymph node, but no evidence of distant metastasis.[42] The potential benefits and risks of deferred surgery for residual primary tumours or metastases after partial response to checkpoint inhibitor treatment is, however, gaining interest, considering the potential for long-lasting effects with these treatments.

2.4.2.2. Systemic treatment

The treatment landscape for RCC has evolved significantly with the introduction of targeted therapies and immunotherapies.

Vascular endothelial growth factor (VEGF) receptor-tyrosine kinase inhibitors (TKIs), encompassing a range of multikinase inhibitors, have emerged as the cornerstone of targeted therapies in the treatment of RCC. These agents target VEGF receptors, primarily 1-3, which play a critical role in tumour-induced angiogenesis and lymphogenesis. Standard treatments for RCC may include various VEGF receptor-TKIs such as sunitinib, pazopanib, tivozanib, and cabozantinib. These inhibitors act by impeding the activity of VEGFRs, thereby disrupting the signalling pathways involved in angiogenesis and lymphogenesis. VEGF receptor-TKIs can be initially classified as selective or non-selective inhibitors. Non-selective inhibitors have the capability to interact with multiple targets and exhibit different levels of *in vitro* potency against VEGF receptors. This potency can range from low (e.g.,

sorafenib) to intermediate (e.g., sunitinib) to high (e.g., cabozantinib and lenvatinib). On the other hand, selective inhibitors demonstrate an increased selectivity for VEGF receptors and display intermediate (e.g., pazopanib) or high (e.g., axitinib, tivozanib) *in vitro* inhibitory activity specifically against VEGF receptors.

In 2015, nivolumab an anti-programmed cell death protein 1 (PD-1) inhibitor was approved for VEGF refractory RCC initiating the rise of immunotherapy in treatment options. The combination of immunotherapy and targeted therapy can achieve higher response rates and better outcomes via additive or synergistic mechanisms. Therefore, various combinations of immunotherapy and targeted therapies have been studied in mRCC. In recent years, antibody-based immunotherapies targeting immune checkpoint receptors PD-1 and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) have demonstrated clinical efficacy in mRCC patients. This led to the approval of nivolumab + ipilimumab as a 1st line approach for mRCC patients with intermediate or poor risk disease.

1st line systemic treatment (untreated aRCC)

In the 1st line treatment of RCC, several options are available depending on the patient's risk profile and individual characteristics. These treatment approaches aim to effectively target and manage the disease while considering factors such as efficacy, tolerability, and patient preferences. Clinical advice to the EAG indicated that when initiating 1st line therapy, the emphasis is on selecting the treatment that offers the best potential for long-term survival. After that the focus shifts more towards palliative measures aimed at managing symptoms and improving HRQoL.

The use of 1st-line PD-1 inhibitor therapy, in combination with VEGF receptor-targeted therapy, has shown improved outcomes compared to TKI monotherapy for patients with clear cell aRCC. This approach harnesses the immune system to fight cancer cells while simultaneously inhibiting the pathways that promote tumour growth and spread. The CLEAR trial[45] evaluated the combination of pembrolizumab + lenvatinib and found that it provided an OS advantage compared to sunitinib alone. The pembrolizumab + lenvatinib combination also showed higher response rates and longer PFS. Pembrolizumab + lenvatinib, along with other TKI + PD-1 inhibitor combinations such as avelumab + axitinib, pembrolizumab + axitinib, or cabozantinib + nivolumab, are licensed as 1st-line treatments for clear cell aRCC, regardless of the patient's risk groups. Pembrolizumab + lenvatinib has recently been recommended by NICE (TA858[38]) in patients who are not eligible for treatment with nivolumab + ipilimumab, while avelumab + axitinib is available via the Cancer Drugs Fund (CDF) (TA645[46]), pembrolizumab + axitinib is not recommended by NICE (TA650[47]), and

cabozantinib + nivolumab is under review by NICE in this appraisal. There is no preferred TKI + PD-1 inhibitor combination in existing guidelines. Although clinical advice to the EAG suggests that pembrolizumab + lenvatinib is likely to be preferred over avelumab + axitinib in intermediate/poor risk patients due to a perceived better efficacy. Clinical advice also indicated that cabozantinib + nivolumab is likely to be considered similar to pembrolizumab + lenvatinib rather than a direct comparator to nivolumab + ipilimumab. One clinical expert considered that the cabozantinib + nivolumab combination may be particularly beneficial for patients with bone metastases due to the cabozantinib component of the treatment.

Nivolumab + ipilimumab is a recommended 1st-line treatment for patients with intermediate- and poor-risk disease (TA780[48]). Clinical advice to the EAG noted that choosing between nivolumab + ipilimumab and pembrolizumab + lenvatinib is challenging in the absence of head-to-head trials. Although nivolumab + ipilimumab is considered to be more toxic, it has more mature survival data available, indicating potential long-term benefits in terms of OS related to its mechanism of action as a combination of immune checkpoint inhibitors. NICE recommendations only allow the use of pembrolizumab + lenvatinib in patients who are not able to take nivolumab + ipilimumab.

For patients who undergo risk stratification and are not eligible for IO therapy, single-agent TKIs such as sunitinib (TA169[49]), pazopanib (TA215[50]), tivozanib (TA512[51]) are alternative treatments, in addition to cabozantinib for those with intermediate- and poor-risk disease (TA542[52]). TKIs work by specifically targetin g the signalling pathways involved in RCC development. While checkpoint inhibitors are generally preferred unless there are strong contraindications, clinical feedback to the EAG indicated the use of 1st line single-agent TKIs is still seen in 30-40% of patients currently. This was considered to be higher than optimal. Evidence from the most recent RWE (UK RWE, 2022[53]) shows 60% of patients were treated with a 1st line single agent TKI in the period 2018 to 2022 (sunitinib 25%, tivozanib 8%, pazopanib 18%, cabozantinib 9%). Although nivolumab + ipilimumab (23.4%) and avelumab + axitinib (12.7%) only became available via CDF from 2019 and 2020 respectively and pembrolizumab + lenvatinib received its recommendation outside of the study period, which may perhaps reflect the high usage of 1st line single agent TKIs in the study period. Of note, ESMO guidelines, consider sunitinib or pazopanib are potential alternatives to PD-1 inhibitor-based combination therapy in IMDC favourable-risk disease due to a lack of clear superiority for PD-1-based combinations over sunitinib in this subgroup of patients.

2nd and subsequent lines of systemic treatment (previously treated aRCC)

The advent of ICI combinations as the standard 1st line therapy for mRCC has raised questions about the best 2nd line treatment strategy in this new treatment landscape. Currently, limited data are available regarding the optimal 2nd line treatment option for patients who have progressed on a 1st line ICI-based combination therapy. International guidelines, such as those from the ESMO,[43] acknowledge the lack of robust prospective data specifically focusing on 2nd line treatment after 1st line PD-1 inhibitor-based combination therapy.

Treatment options for 2nd line therapy could include a TKI, a PD-1 inhibitor or a mammalian target of rapamycin (mTOR) mTOR inhibitor. Immune checkpoint inhibitors cannot be given more than once in the systemic treatment pathway and therefore nivolumab is not an option. It is also reasonable to consider using a TKI that was not utilised in the 1st line combination as a potential 2nd line treatment option, as there are reasonable probabilities of achieving further clinical benefit with this approach.

In patients who were initially treated with the combination of immunotherapy and VEGF receptor-targeted therapy (e.g. avelumab + axitinib, pembrolizumab + lenvatinib), treatment options in the 2nd line include axitinib,[54] cabozantinib,[55] lenvatinib + everolimus,[56] and everolimus (TA432[57]) depending on the 1st line treatment combination received:

- avelumab + axitinib (TA645[46]) → cabozantinib (TA463[55]), lenvatinib + everolimus (TA432[56]), or everolimus (TA432[57]).
- pembrolizumab + lenvatinib (TA858[38]) → axitinib (TA333[54]), cabozantinib (TA463[55]), or everolimus (TA432[57]).

While the majority of patients receive cabozantinib, in certain cases lenvatinib + everolimus may be considered as an alternative as it as it can only be used after one prior TKI. This option may be preferred in an effort to maximise the available lines of treatment for patients. Clinical advice indicated that lenvatinib + everolimus is preferred over everolimus monotherapy as it allows for a lower dose of everolimus and improved tolerability. Axitinib is not commonly used as a 2nd line treatment and is often reserved for later lines of therapy. Otherwise, 1st line options of sunitinib (still on label as 2nd line treatment) or pazopanib (off label as 2nd line treatment), or tivozanib (off-label as 2nd line treatment) may also be considered. Clinical feedback to the EAG anticipated that following cabozantinib + nivolumab, lenvatinib + everolimus is likely to be preferred as it provides a different approach to the previous regimen.

In patients who were initially treated with the combination of nivolumab + ipilimumab, the treatment options after disease progression include cabozantinib, sunitinib (still on label as 2nd line treatment), pazopanib (off label as 2nd line treatment), or tivozanib (off-label as 2nd line treatment). Clinical advice to the EAG indicated that cabozantinib is typically chosen as the next treatment option (although the EAG note that it is off-label following nivolumab + ipilimumab), as administering another round of checkpoint inhibitor therapy is generally considered futile and is also not allowed in the UK.

In patients who were initially treated with VEGF receptor-directed TKI monotherapy, the recommended treatment options after disease progression include nivolumab (TA417[58]) or cabozantinib (TA463[55]), both of which demonstrated OS benefit in the 2nd-line setting. Other options that can be considered include axitinib (TA333[54]), and lenvatinib + everolimus (TA498[56]).

While approved for 2nd line and 3rd line treatment, clinical advice to the EAG indicated that everolimus and axitinib are typically reserved for 4th line treatment. Although given the toxicity of everolimus only a small proportion of patients would be eligible to receive it due to toxicity.

In England recommendations for subsequent treatments are provided in the Cancer Drugs Fund (CDF) list:[37] These broadly reflect the above. The CDF list only includes drug indications which became available from 2016 onwards when the BlueTeq® high-cost drug management system started to be routinely implemented. Sunitinib, pazopanib and axitinib were therefore not included. Information provided on recommended subsequent therapies, follow-up and treatment breaks has also become more detailed over time. The CDF recommendations demonstrate that 1st line TKIs are recommended and available in the 2nd line setting in the NHS, with two of these being used off-label, as shown in Figure 6.

In clinical practice, expert advice suggested that it is realistic to expect that most patients with RCC would receive up to three lines of treatment. Approximately 10-20% of patients may reach the 4th line of treatment. However, it is uncommon for patients to go beyond the 4th line, and very few would require a 5th line of treatment. This is in line with the UK RWE dataset identified for this pilot.[53] The introduction of new therapies, such as belzutifan, may change the treatment landscape and potentially replace everolimus as a last-line treatment option.

2.4.2.3. Best supportive care

For individuals who cannot tolerate or do not wish to receive active treatment, best supportive care (BSC) is provided. BSC focuses on monitoring disease progression, symptom control, and palliative care without active treatment.[50]

The treatment pathway overview is summarised in Figure 6 and possible treatment sequences are summarised in Figure 7.

Table 1: Current treatment options per NICE recommendations for untreated aRCC

Intervention	Suni	Pazo	Cabo	Tivo	Nivo+ipi	Ave+axi	Pem+lenv
NICE appraisal	TA169[49]	TA215[50]	TA542[52]	TA512[51]	TA780 (CDF review of TA581)) [48]	TA645 (CDF)[46]	TA858[38]
Class	TKI	TKI	TKI	TKI	PD-1 inhibitor + CTLA-4 inhibitor	PD-1 inhibitor + TKI	PD-1/ PD-L1 inhibitor + TKI
Recommendation	1L (ECOG PS 0 or 1)	1L (no prior cytokine therapy; ECOG PS 0 or 1)	1L	1L	1L	1L via CDF	1L (if not suitable for nivo+ipi)
IMDC risk category	All risk	All risk	Int or poor risk per IMDC criteria	Int or poor risk per IMDC criteria	Int or poor risk per IMDC criteria	All risk	Int or poor risk per IMDC criteria

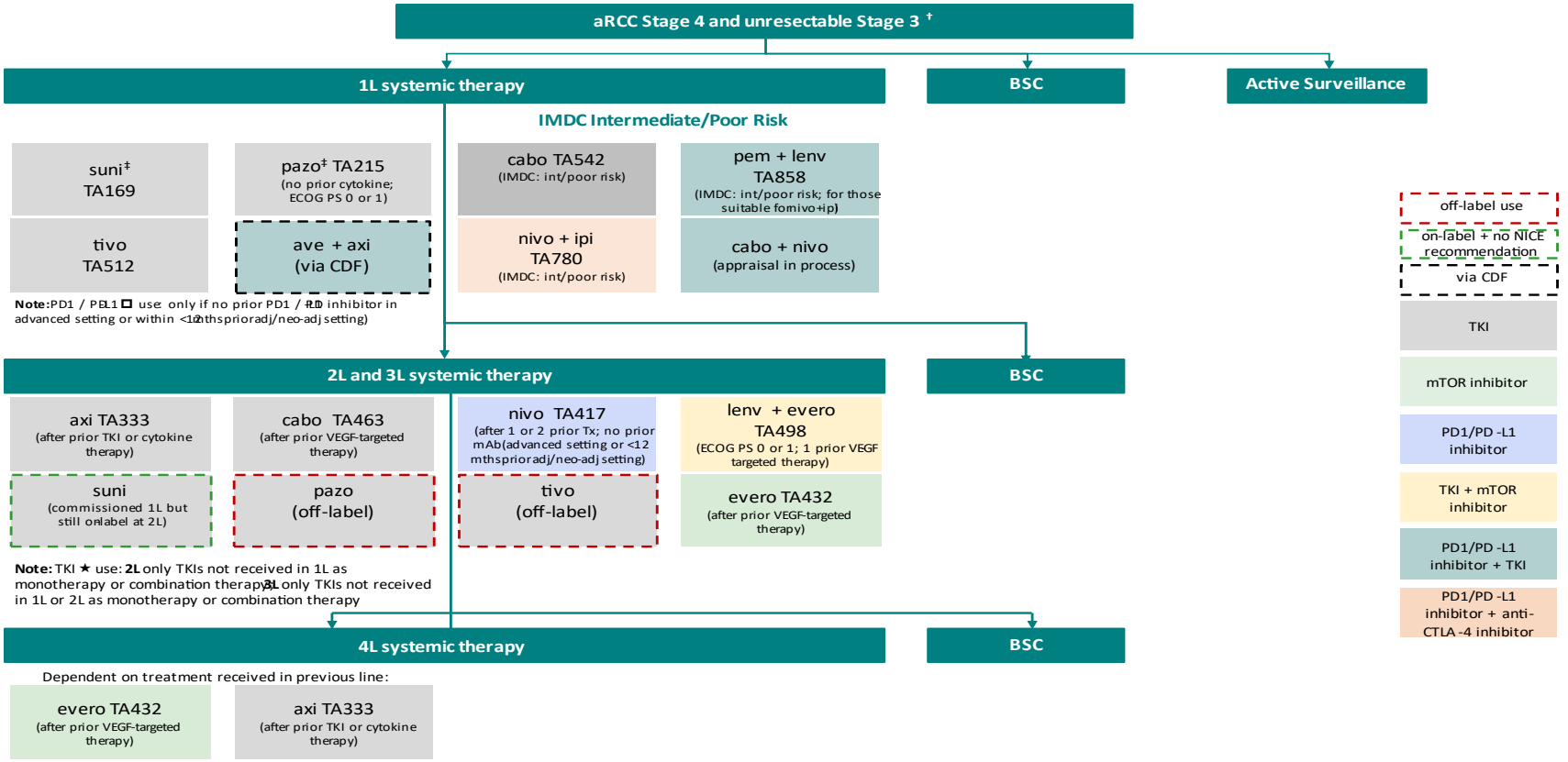
Abbreviations: 1L, 1st line; aRCC, advanced renal cell carcinoma; ave, avelumab; axi, axitinib; cabo, cabozantinib; CDF, Cancer Drugs Fund; CTLA-4, cytotoxic T-lymphocyte-associated protein 4; IMDC, International Metastatic renal cell carcinoma Database Consortium; ipi, ipilimumab; lenv, lenvatinib; NICE, National Institute for Health and Care Excellence; nivo, nivolumab; pazo, pazopanib; PD1, programmed death 1; PD-L1, progress death cell ligan 1; pem, pembrolizumab; suni, sunitinib; TA, technology appraisal; tivo, tivozanib; TKI, tyrosine kinase inhibitor

Table 2: Current treatment options per NICE recommendations for untreated aRCC

Intervention	Evero	Axi	Nivo	Cabo	Lenv+evero
NICE appraisal	TA219 → TA432[57]	TA333[54]	TA417[58]	TA463[55]	TA498[56]
Class	mTOR inhibitor	TKI	PD-1	TKI	TKI + mTOR inhibitor
Line of treatment recommended	2L (after prior VEGF-targeted therapy)	2L (after 1L TKI or a cytokine [not recommended by NICE])	2L (after 1 or 2 prior treatments; no prior mAb (advanced setting or <12 mths prior adj/neo-adj setting)	2L (after prior VEGF-targeted therapy)	2L (after 1 prior VEGF targeted therapy and ECOG PS 0 or 1)
Sequencing notes	Clinical advice to the EAG notes that evero would primarily be used at 4L	Clinical advice indicates axi would not be used after tivo as they have a similar MoA			

Abbreviations: 2L, 2nd line; axi, axitinib; adj, adjuvant; cabo, cabozantinib; ECOG, Eastern Cooperative Oncology Group performance status; evero, everolimus, lenv, lenvatinib; mAb, monoclonal antibody; MoA, mechanism of action; mths, months; mTOR, mammalian target of rapamycin; NICE, National Institute for Health and Care Excellence; nivo, nivolumab; PD-1, programmed death 1; PS, performance status; TA, technology appraisal, tivo, tivozanib; TKI, tyrosine kinase inhibitor; VEGF, vascular endothelial growth factor receptor

Figure 6: Treatment pathway for advanced stage RCC: overview



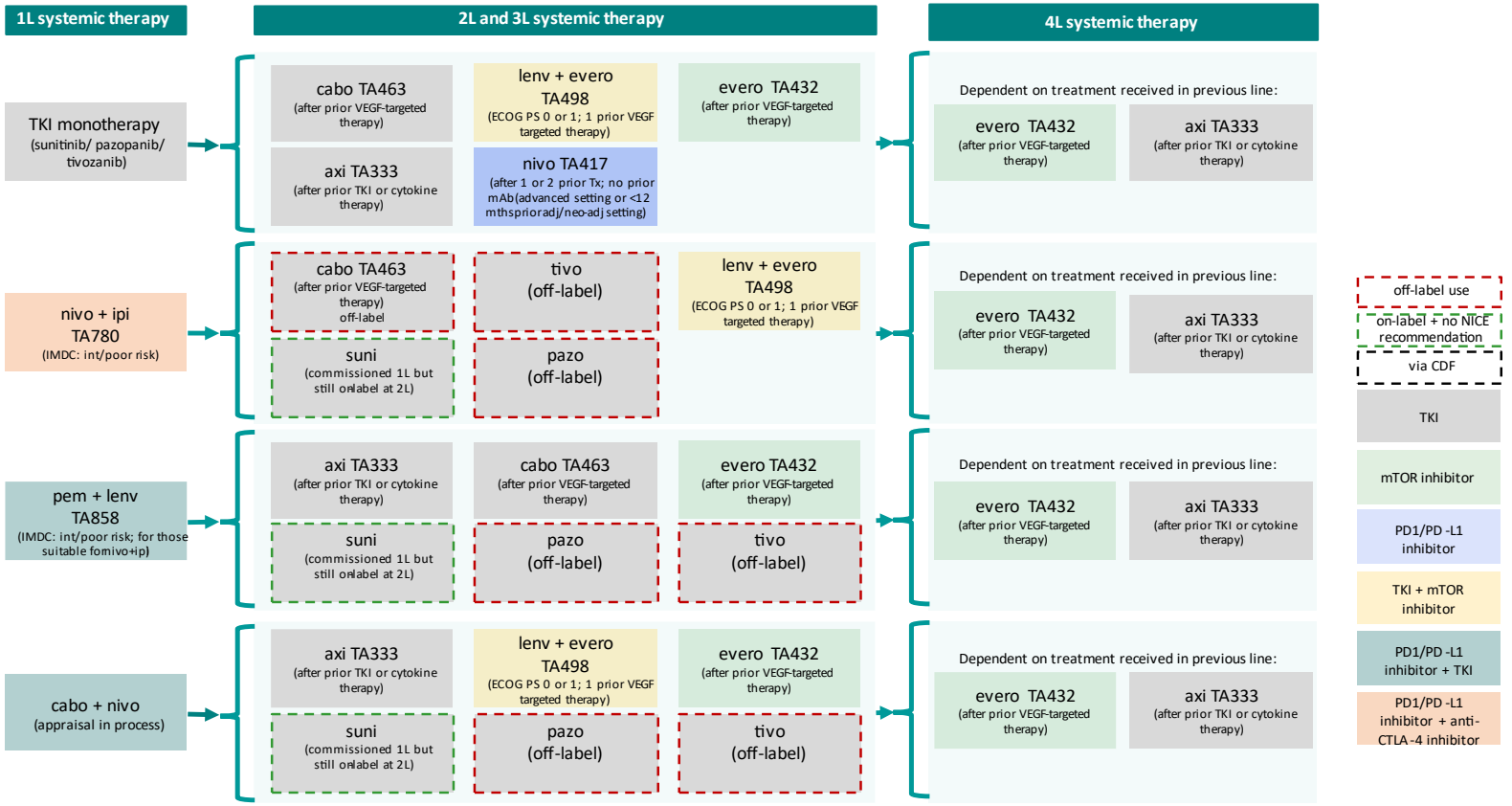
Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; adj adjuvant; aRCC, advanced renal cell carcinoma; ave, avelumab; axi, axitinib; BSC, best supportive care; cabo, cabozantinib; CDF, Cancer Drugs Fund; ESMO, European Society of Medical Oncology; evero, everolimus; IMDC, International Metastatic renal cell carcinoma Database Consortium; int, intermediate; ipi, ipilimumab; lenv, lenvatinib; mTOR, mammalian target of rapamycin; nivo, nivolumab; pazo, pazopanib; PD1, programmed cell death protein 1; PD-L1, programmed death ligand 1; pem, pembrolizumab; suni, sunitinib; TA, technology appraisal; tivo, tivozanib; TKI, tyrosine kinase inhibitor; Tx, treatment; VEGF, vascular endothelial growth factor

Notes:

†Cancer has spread into surrounding tissues outside Gerota's fascia or into adrenal gland. Cancer has spread to another part of the body. May or may not spread to lymph nodes

‡Also considered potential alternatives to PD-1 inhibitor-based combination therapy in IMDC favourable-risk disease (ESMO guideline recommendations; 2021)

Figure 7: Treatment pathway for advanced stage RCC: possible treatment sequences



Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; adj adjuvant; aRCC, advanced renal cell carcinoma; ave, avelumab; axi, axitinib; BSC, best supportive care; cabo, cabozantinib; CDF, Cancer Drugs Fund; ESMO, European Society of Medical Oncology; evero, everolimus; IMDC, International Metastatic renal cell carcinoma Database Consortium; int, intermediate; ipi, ipilimumab; lenv, lenvatinib; mTOR, mammalian target of rapamycin; nivo, nivolumab; pazo, pazopanib; PD1, programmed cell death protein 1; PD-L1, programmed death ligand 1; pem, pembrolizumab; suni, sunitinib; TA, technology appraisal; tivo, tivozanib; TKI, tyrosine kinase inhibitor; Tx, treatment; VEGF, vascular endothelial growth factor

Notes:

†Cancer has spread into surrounding tissues outside Gerota's fascia or into adrenal gland. Cancer has spread to another part of the body. May or may not spread to lymph

nodes
‡Also considered potential alternatives to PD-1 inhibitor-based combination therapy in IMDC favourable-risk disease (ESMO guideline recommendations; 2021)

2.5. Decision problem

As noted in Section 1, this pilot is designed to address a broader decision problem than is considered within a standard STA. The platform model to be developed encompasses all stages of the treatment pathway for RCC, including all treatments within the treatment pathway for 1st and subsequent line systemic treatment (Section 2.4). Within the pilot and summarised in this report, the EAG appraised the clinical and cost effectiveness of one new treatment: cabozantinib + nivolumab for untreated advanced or metastatic RCC. A summary of the decision problem for the appraisal of this treatment is provided in Table 3.

Table 3: Summary of decision problem

	Final scope issued by NICE	Decision problem addressed
Population	People with untreated advanced or metastatic RCC	Per the scope, all evidence identified was for adults
Intervention	Cabo+nivo (submission led by Ipsen)	Per the scope
Comparator(s)	• Pazo • Tivo • Suni • Cabo (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Nivo+ipi (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Pem+lenv (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Active surveillance	In line with the scope except that active surveillance has not been included as it is considered to happen prior to the decision node at which this model starts. Clinical advice received is that clinical decision-making first involves deciding whether a person would benefit from any kind of systemic therapy and then, once the decision to initiate therapy has been taken, a choice is made between available treatment options
Outcomes	• Overall survival (OS) • Progression-free survival (PFS) • Response rates • Duration of response • Time on treatment/time to next treatment (TTND) • Adverse effects of treatment • Health-related quality of life (HRQoL)	Per the scope dependent upon data availability; limited data are available for time on treatment and time to next treatment within published literature

	Final scope issued by NICE	Decision problem addressed
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any commercial arrangements for the intervention, comparator or subsequent treatment technologies will be taken into account.	Per the scope
Subgroups	If the evidence allows the following subgroup will be considered: Intermediate-/poor-risk advanced metastatic RCC as defined in the IMDC criteria Prior treatment	Per the scope. Data are not available within CheckMate 9ER to explore the impact of prior adjuvant treatment on outcomes
Special considerations including issues related to equity or equality	None	None

Abbreviations: IMDC, International Metastatic RCC Database Consortium; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; RCC, renal cell carcinoma

2.6. Description of the technology being evaluated

Cabozantinib is a multiple receptor TKI and nivolumab is a PD-1 inhibitor. The combination was granted approval for the 1st line treatment of advanced RCC on the basis of the CheckMate 9ER Phase 3 trial[59], first by the European Medicines Agency (EMA) on 26th March 2021[60] and then Medicines and Healthcare products Regulatory Agency (MHRA) on 13th May 2021.[61] The marketing authorisation holder for cabozantinib is Ipsen Pharma. The marketing authorisation holder for nivolumab is Bristol-Myers Squibb Pharma EEIG.

Cabozantinib is administered orally at a dose of 40 mg once daily.[62] Nivolumab is given intravenously at a dose of either 240 mg every two weeks or 480mg every four weeks: the former was used in CheckMate 9ER while, based upon initial expert consultation, the latter is more likely to be used in clinical practice. In line with the trial, the Summary of Product Characteristics (SmPC)[62] specifies that cabozantinib *"should be continued until disease progression or unacceptable toxicity. Nivolumab should be continued until disease progression, unacceptable toxicity, or up to 24 months in patients without disease progression."*

Table 4. Technology being evaluated

UK approved name and brand name	**Cabo+nivo**
Mechanism of action	Receptor tyrosine kinases (RTKs) are receptors for many growth factors and proteins implicated in the development and progression of cancer, including [63-65]: • Vascular endothelial growth factor (VEGF) which promotes the growth of new blood vessels • Hepatocyte growth factor that regulates several physiological processes including proliferation, scattering, morphogenesis, and survival of cells, and • Growth factor growth arrest specific 6 (GAS6) which is involved in several cellular functions including growth, migration, aggregation, and differentiation Cabo is a tyrosine kinase inhibitor (TKI) that inhibits multiple RTKs involved in tumour growth and angiogenesis, pathologic bone remodelling, drug resistance, and metastatic progression of cancer [14]. Cabo is a potent inhibitor of multiple RTKs, such as c-MET and VEGF, known to play important roles in tumour cell proliferation and/or tumour neovascularisation in RCC [66,67]. There is an interaction between angiogenesis and immunosuppression in tumour development. VEGF primarily inhibits the innate immune system by upregulating PD-L1 and CTLA-4 expression, thereby maintaining an immunosuppressive environment. In addition, antiangiogenic activity leads to normalisation of the tumour vasculature and exhibits a positive effect on immune-cell infiltration into tumours [68].

UK approved name and brand name	**Cabo+nivo**
	Nivo is a fully human, monoclonal immunoglobulin antibody (IgG4) that acts as a checkpoint inhibitor of PD-1 and blocks its interaction with its ligands. Tumours use PD-L1 expression as defence or escape mechanisms against the host's anti-tumour T cell response; inhibiting PD-L1 restores the function of these anti-tumour T cells which have become ineffective or suppressed [68]. Therefore, the efficacy of PD-L1 inhibition relies on a pre-existing immune response [68]. The combination of cabo+nivo therefore potentiates immune-mediated tumour destruction in parallel to targeted inhibition of tumour growth and progression.
Marketing authorisation/ CE mark status	Cabo+nivo received MHRA approval on 13/05/2021.[69]
Indications and any restriction(s) as described in the SmPC	In accordance with the current marketing authorisation, cabo+nivo is indicated for the treatment of previously untreated adult patients with aRCC or mRCC. Cabometyx® (cabo) monotherapy is licensed for the following indications[69]: • Treating aRCC in treatment-naïve adults with intermediate or poor-risk • Treating aRCC in adults following prior VEGF-targeted therapy • Treating hepatocellular carcinoma in adults who have previously been treated with sora • Treating adults with locally advanced or metastatic DTC, refractory or not eligible to radioactive iodine who have progressed during or after prior systemic therapy. Opdivo® (nivo) monotherapy is licensed for the following indications[70]: • Treating aRCC after prior therapy in adults • Treating advanced (unresectable or metastatic) melanoma in adults • Adjuvant treatment of adults with melanoma with involvement of lymph nodes or metastatic disease after complete resection • Treating locally advanced or metastatic NSCLC after prior chemotherapy in adults • Treating adult patients with relapsed or refractory classical Hodgkin's lymphoma after autologous stem cell transplantation and treatment with brentuximab vedotin • Treating recurrent or metastatic squamous cell cancer of the head and neck in adults progressing on or after platinum-based therapy • Treating locally advanced, unresectable, or metastatic urothelial carcinoma in adults after failure of prior platinum-containing therapy
Method of administration and dosage	Cabo is available as 20 mg, 40 mg, and 60 mg film-coated tablets. The recommended dose for cabo is 40 mg once daily in combination with nivo 240 mg every two weeks or 480 mg every 4 weeks. The treatment should continue until disease progression or unacceptable toxicity. Nivo treatment should continue until disease progression or

UK approved name and brand name	Cabo+nivo
	unacceptable toxicity or up to a maximum duration of 2 years in patients without disease progression[69,70]. For cabo, temporary treatment interruption and/or dose reduction is recommended for management of adverse drug reactions. In monotherapy, dose is reduced to 40 mg daily, and further to 20 mg daily. Whereas, in combination with nivo, it is recommended to reduce the dose to 20 mg of cabo once daily, and then to 20 mg every other day. For nivo, dose reduction is not recommended, and in case of AEs or liver enzymes elevation, either withhold dose or discontinue treatment [69,70].
Additional tests or investigations	No additional tests or investigations are needed to identify patients eligible for cabo+nivo over those needed to identify advanced or mRCC.
List price and average cost of a course treatment	List price: £5,143.00 per 30 x 40 mg tablet pack of cabo[71] £1,097.00 per 100 mg vial; £439.00 per 40 mg vial of nivo[72]
Patient access scheme (if applicable)	A confidential simple patient access scheme is available for cabo. The pack price under this scheme is [redacted] (a [redacted]% discount to the list price). There is a confidential patient access scheme in place for nivo, approved by the DHSC.

Abbreviations: AEs, adverse events; aRCC, advanced RCC; DHSC, Department of Health and Social Care; DTC, differentiated thyroid cancer; cabo, cabozantinib; DHSC, department of health and social care; MHRA, Medicines and Healthcare Products Regulatory Agency; mRCC, metastatic renal cell carcinoma; nivo, nivolumab; NSCLC, non-small-cell lung cancer; PD-1 programmed death 1; PD-L1, programmed death-ligand 1; SmPC, summary of product characteristics; sora, sorafenib; VEGF, vascular endothelial growth factor receptor

Notes: information taken from company submission

2.7. Equality considerations

No equality issues were identified within this appraisal.

3. METHODS FOR REVIEWING CLINICAL EFFECTIVENESS

3.1. Assessment group methods for reviewing clinical evidence

The EAG conducted a systematic literature review (SLR) to identify published evidence and real-world data sets relevant to the decision problem. The methods used were consistent with the NICE preferred methods and with best practice guidance for the conduct of SLRs.[73,74] This section provides:

- A description of the methods used to identify published RCT evidence;
- A description of the methods used to identify real-world data;
- A summary of the methods used to gather clinical input; and
- Information on how data from the company submission was considered

3.1.1. Identification of systematic literature reviews and randomised controlled trials

3.1.1.1. Search strategies and screening process

Systematic searches were conducted to identify 1) clinical effectiveness SLRs and meta-analyses, and 2) randomised controlled trials (RCTs) published since the most recent relevant systematic reviews. The database searches were complemented by supplemental searching, such as citation chasing and hand-searches of grey literature sources. All data from published HTA reports included in the reviews were publicly available; i.e. redacted data from published NICE HTA reports were not included.

Search strategies were developed by an information specialist and quality assured by another information specialist. The search strategies used a combination of indexed keywords (e.g., Medical Subject Headings [MeSH]) and free-text terms appearing in the titles and/or abstracts of database records and were adapted according to the configuration of each database. No limits on publication status (published, unpublished, in-press, and in-progress) were applied. The strategy used for SLR and RCT evidence is described in the following sections. The searches from NICE TA858[38] were used as a starting point for development of search terms for this appraisal. Full search strategies are supplied in Appendix A.

Articles for the SLR and RCT searches were independently assessed for inclusion by two reviewers using the pre-specified inclusion/exclusion criteria. Discrepancies were resolved by

discussion, with involvement of a third reviewer, where necessary. All duplicate papers were double checked and excluded.

Search for RCTs within published SLRs

Searches for relevant SLRs were undertaken in MEDLINE, Embase, Cochrane Database of Systematic Reviews (CDSR) and The International Network of Agencies for Health Technology Assessment (INAHTA) up to 19th December 2022. Relevant NICE technology appraisals were identified by handsearching the NICE website and were screened for further relevant studies.

The search used a combination of terms for RCC with relevant intervention terms. There were no restrictions on cancer stage or line of treatment for this search. The intervention terms were avelumab, axitinib, cabozantinib, everolimus, ipilimumab, lenvatinib, nivolumab, pazopanib, pembrolizumab, sunitinib, and tivozanib, plus relevant brand names and other alternative names.

In MEDLINE and Embase, the EAG used the systematic review, meta-analysis and HTA filter from The Canadian Agency for Drugs and Technologies in Health (CADTH)[75] to identify relevant records. All searches were limited from 2018 onwards; however, as the searches resulted in a high volume of hits (n=1,273 after de-duplication), a decision was taken to limit screening to records published from 2020 onwards (thereby excluding 371 retrieved records published pre-2020). No language filters were used. Conference abstracts were included.

The most recent, highest-quality and most comprehensive SLRs were then sought to identify RCTs relevant for this appraisal. The SLRs identified were qualitatively assessed against the following criteria:

- Is a full paper available (rather than an abstract)?
- Which line(s) of treatment were included?
- How many treatments specified within the decision problem were included within the networks?
- Were the trials included in the most recent NICE TAs for the relevant line of treatment included (TA858, TA645, TA463)?
- For SLRs looking at 1st line treatments: were data presented by risk subgroup?
- Does the methods description indicate that this is a high quality SLR?

Based upon these criteria, four SLRs were identified and screened for RCTs: Heo 2022, Liao 2022, Riaz 2021 and NICE TA858.[38,76-78] The publication date of these SLRs was then used to inform the date from which to run the top-up RCT searches described in the next section.

Heo et al. presented a SLR and network meta-analysis (NMA) of OS and PFS for 1st and 2nd line therapies in participants with advanced RCC based upon 26 RCTs (1st line: 19; 2nd line: 9) with 13,893 participants. The networks presented included a number of treatments that are not available in the NHS, and the search excluded three treatments of interest to our decision problem: cabozantinib + nivolumab, pembrolizumab + lenvatinib in 1st, and everolimus + lenvatinib for people who have been previously treated. The authors searched for trials published between 2000 and June 2020 which would be expected to capture all trials for treatments included in the decision problem for this appraisal given when development of the relevant treatments began. The review was conducted using best practice methods.

Liao et al. presented a SLR and NMA for advanced RCC treatments in the 2nd line setting. Nine RCTs with 4,911 participants were included. The study considered all systemic treatments used in a 2nd line setting and therefore identified evidence for everolimus + lenvatinib, which was missing from the Heo et al. study. Searches were conducted from inception to 20th July 2021. The study reporting was less comprehensive than Heo et al.; however, the study was included due to the broader range of treatments covered and more recent search date.

Riaz et al. presented a living, interactive SLR and NMA of 1st line treatments for advanced RCC. No limits on included treatments were imposed and outcomes were presented by risk score. Evidence was identified for all of the 1st line RCC treatments of interest to the decision problem, except for pembrolizumab + lenvatinib. A comprehensive search was conducted from inception by an experienced medical librarian in consultation with the principal investigator (I.B.R.). A “living” auto search with monthly updates was subsequently created with the last date of evidence included being 22nd October 2020. Study selection and extraction were both semi-automated.

TA858 was the most recent NICE TA in RCC. This appraisal considered treatments in the 1st line setting, and searches were run in October and November 2021. All the 1st line treatments of interest were included with the exceptions of avelumab + axitinib and cabozantinib + nivolumab. Reporting was split by risk group. Screening and extraction were performed by two reviewers.

Full search strategies were provided in the appendix of the report for TA858 and were used to inform the development of the searches conducted within this appraisal.

Top-up search for additional RCTs

A top-up search to identify RCTs published since the latest SLR search dates was conducted. The search was conducted in MEDLINE, Embase, CENTRAL (Cochrane Central Register of Controlled Trials) and trial registers (WHO International Clinical Trials Registry Platform and Clinicaltrials.gov). The search identified trials published from 2021 onwards, which allowed a reasonable overlap in time to capture RCTs published since the most recent search dates of the reviews for each line of treatment: Liao 2022 and TA858.[38,77]

RCTs were identified using the same intervention terms as used in the search for SLRs. For this search terms focusing on people with advanced, metastatic or otherwise later stage RCC were used. The Cochrane RCT filter was used to identify relevant trials in MEDLINE and Embase. No language limits were applied. Conference abstracts were included.

Scopus was searched for subsequent data cuts of trials included in the identified SLRs, including conference abstracts. Further citation searches were conducted (forward and backward citation searching) in Scopus for all additional RCTs identified that were not included in the latest SLRs. Relevant NICE technology appraisal reports were reviewed to identify any additional unredacted data that had not been subsequently published. The list of published abstracts from the American Society of Clinical Oncology Genitourinary Cancers Symposium, held in San Francisco on the 16 - 18 Feb 2023 (ASCO GU 2023) and the American Society of Clinical Oncology Annual meeting, held in Chicago on the 2-6 June 2023, were hand searched, to identify new trials or new data cuts of already identified trials.

Finally, HRQoL and patient-reported outcomes for the 30 included RCT studies were identified by reviewing the economic searches for the development of the cost-effectiveness model (see Section 4.1.1). Twenty-nine potentially relevant reports were identified by searching for RCT trial numbers in the economic studies Endnote database, which were then sifted down to 23 studies (covering 16 of the 30 RCTs) during full-text review.

To identify ongoing RCTs, the EAG searched Clinicaltrials.gov and WHO International Clinical Trials Registry Platform (ICTRP). The advanced search functionality was used for both platforms, using a combination of intervention terms, population terms, and keywords to identify

RCTs (“random” or “randomized” or “randomised” or “randomisation” or “randomization” or “RCT”). No date or recruitment status limits were applied. The RCT update search of Cochrane CENTRAL (described above) also retrieved registry records.

Contact with study authors

Where data were missing in the published clinical effectiveness studies, the EAG wrote to the authors. This was only done where data for an entire key outcome, Kaplan-Meier data for a key outcome or subgroup data (baseline characteristics or outcomes) were missing. A deadline for response to the initial contact of four weeks was imposed. Additional time was allowed where the author indicated they were able to supply the data requested and where it did not impact on the broader timelines for this appraisal. No responses were received via this route which could be included as agreement was required from the companies funding the relevant trials. Additional data was received for CheckMate 214 from BMS within their response to the preliminary assessment report.

3.1.1.2. Inclusion and exclusion criteria

In the first round of screening, SLRs that included RCTs of pharmacological treatments for advanced RCC published since 2020 were included. Reviews focusing on the efficacy of radiotherapy or surgical interventions were excluded. The highest-quality and broadest systematic reviews were then used to identify relevant RCTs, from which line of treatment and comparators were extracted and compared to the full platform model decision problem to identify any gaps.

In top-up searches, RCTs for people with advanced RCC of systemic treatments funded within the NHS (pazopanib, tivozanib, sunitinib, cabozantinib, nivolumab + ipilimumab, lenvatinib + pembrolizumab, axitinib, lenvatinib + everolimus, everolimus, cabozantinib + nivolumab, nivolumab, avelumab + axitinib, best supportive care) were included where they reported at least one outcome from OS, PFS, TTNT, TTD, response rates, adverse effects of treatment, and HRQoL. As a protocol clarification, the EAG also included studies with placebo as a comparator and only included studies with relevant comparisons of drugs prescribed at the licensed doses. In addition, as a protocol deviation, the EAG included studies with sorafenib as a comparator. This is because past TAs have acknowledged the importance of sorafenib as a linking treatment in evidence networks and the EAG also anticipated needing to use sorafenib as a linking treatment.

Further details on these inclusion/exclusion criteria used for SLRs and RCTs are presented in Table 5.

Table 5: Inclusion and exclusion criteria: SLRs and RCTs

	Include	Exclude
Population	Studies of participants with advanced (unresectable Stage 3 or Stage 4) RCC at any treatment line	Studies of participants with early stage (not advanced) RCC
Intervention	Round 1 (systematic reviews): any pharmacological treatment for advanced RCC used in the systemic setting Round 2 (RCTs and extensions of RCTs): cabo+nivo, pazo, tivo, suni, cabo, nivo+ipi, pem+lenv, axi, lenv+evero, evero, nivo, ave+axi*, Sora and placebo were included as linking treatments for use in the NMA	Any other treatments not listed under inclusion Treatments used in the adjuvant setting
Comparator	• Any of the other interventions listed above (i.e. head-to-head studies) • Dose comparison studies • Usual care / physicians' choice / BSC / placebo	Non-pharmacological treatments only
Outcomes	Studies reporting at least one outcome from: • OS • PFS • TTNT • Time on treatment • Response rates • Duration of response • AEs of treatment† • HRQoL	Studies not reporting an included outcome
Study design	Round 1: systematic reviews of RCTs published since 2020 Round 2: RCTs. The most recent conference abstract for each intervention and outcome will be included unless a full journal article is available	Round 1: systematic reviews that did not contain RCTs, systematic reviews of treatment effect modifiers. Round 2: non-randomised trials, observational studies, case reports, editorials and commentaries

Abbreviations: AE, adverse events; BSC, best supportive care; HRQoL, health-related quality of life; NMA, network meta-analysis; OS, overall survival, PFS, progression-free survival; RCC, renal cell carcinoma; RCT, randomised controlled trials; TEAEs, treatment-emergent adverse events; TTNT, time to next treatment

Notes: * As belzutifan was included within the NICE draft scope it was included within the search terms for the searches conducted, these studies will, however, not be included during screening † Grade 3+ TEAEs and the total number of treatment-emergent adverse events leading to discontinuation will be extracted. Additional lower grade AEs of interest may be extracted following clinical advice

3.1.1.3. Data extraction and quality assessment strategy

All relevant published evidence for a given trial is extracted in one single entry in the data extraction matrix. Included clinical effectiveness studies (identified via SLRs and top-up searches) were extracted by one reviewer into a bespoke database and checked by a second reviewer. The data extraction grid is provided in Appendix D. Discrepancies were resolved by discussion, with the involvement of a third reviewer where necessary. For time to event outcomes, both summary hazard ratios and figures for Kaplan Meier curves from the last data cut were extracted. Digitisation of curves using standard methods (the Guyot algorithm[79]) was conducted, assuming censoring linearly across time intervals.

Quality assessments of individual studies were assessed by one reviewer in Microsoft Excel and checked by a second reviewer. Any disagreements were resolved through discussion, with arbitration by a third reviewer if consensus could not be reached. RCTs were assessed using standardised criteria for critically appraising the quality of clinical effectiveness evidence as recommended by NICE for submissions to its HTA programme.[80] The assessment included the consideration of domains that could pose a variable risk of bias for individual outcomes at the outcome level (performance and detection bias, attrition bias, and reporting bias), and identifying any other sources of bias resulting from a design or methodological feature of the study. The latter included bias considerations specific to trial designs that include an element of treatment switching (i.e., crossover trials assigning sequential treatments as well as trials allowing crossover following disease progression) as such trials are prone to carryover bias in the period following the switch due to residual treatment effect from the previous period.

A determination of overall domain bias was made based on the worst-rated of the sub-domains – for example, overall selection bias would be determined by the worst-rated of the randomisation, allocation concealment and baseline imbalance domains. A determination of overall study bias was additionally assessed by considering the key domains for parallel RCTs (selection and attrition bias) and crossover RCTs (selection, attrition and other bias); the overall judgment represented the worst-rated of these domains. Performance and detection biases were omitted from key domains for overall bias considerations as primary outcomes in cancer trials are predominantly hard, objective outcomes; reporting bias was similarly omitted as a key domain as the primary outcomes that inform sample size calculations are rarely omitted from reported results. Finally, biases related to conflict of interest were also omitted as a key domain

since these conflicts are usually present in cancer trials due to manufacturer sponsorship, but influences are carefully monitored and managed in such trials.

It is important to note that the approach to quality assessment in this report is different to that taken in TA858; with the latter being informed by the Centre for Review and Dissemination (CRD) Guidance for Undertaking Reviews in Healthcare. The EAG's approach, following NICE guidance, specifically evaluated the *adequacy* of methods to minimise bias, rather than evaluating whether such methods were reportedly followed (e.g., NICE guidance calls for the assessment of 'Was the allocation adequately concealed?', while in TA858 the question under consideration was 'Was the allocation of treatment concealed?').

3.1.2. Identification of real-world evidence

3.1.2.1. Searches for real-world evidence

In line with the recommendations in the NICE RWE framework,[81] a systematic search process was followed to identify real-world (observational) evidence to characterise the treatment pathway, the natural history of the disease and the characteristics of people with RCC treated in clinical practice. A four-pronged search strategy was used:

1. MEDLINE and Embase: Search results for observational studies in the UK about RCC were uploaded into Endnote, followed by assessment of abstracts to identify any registry/RWE data sources used. The search combined the Scottish Intercollegiate Guidelines Network (SIGN) observational studies filter[82] and the NICE UK filter.[83] Search strategies are provided in Appendix A. Results (n = 2,683) were exported into Endnote and screened by one reviewer using the pre-specified inclusion criteria (Section 3.1.2.3).
2. Health Data Research UK Innovation Gateway: Search terms included 'renal cell cancer', 'renal cell carcinoma', 'kidney cancer' or 'kidney carcinoma'. Results were sifted on screen by one review using the inclusion criteria.
3. Web search (Google and Bing): Individual searches within each database were conducted using terms for RCC and RWE. RCC search terms were: 'renal cell cancer', 'renal cell carcinoma', 'kidney cancer', and 'kidney carcinoma. RWE search terms were 'registry', 'real-world data', and 'real-world evidence'. The first 50 results of each search were sifted on screen by one reviewer using the inclusion criteria.
4. Reviewers flagged potential evidence sources—that met the inclusion—during screening of the main clinical and economic search results.

Further to the above-described search process, RWE sources were also identified from company and stakeholder submissions during the research process. Table 25 describes the potential sources of RWE found and from where they were identified.

3.1.2.2. Screening process

Articles identified from the RWE searches were assessed in a targeted manner by one reviewer using the pre-specified inclusion/exclusion criteria (see Section 3.1.2.3). The potential uses for this evidence are listed below. In each case information was considered for both the whole patient population and according to IMDC risk score subgroups:

- Understand current treatment pathways (sequences) being used.
- Assess the generalisability of trial data based on demographic and disease-related characteristics (particularly prognostic variables).
- Improve long-term extrapolations (particularly for historical therapies).
- Inform baseline risk (either as scenario analysis or base case).
- Understand the difference between trial-based assessment of progression and intermediate disease-related outcomes recording in practice.
- Inform doses used in practice for treatments where dose adjustments can be applied & understand the proportion of planned doses that are missed.
- Look at how HRQoL changes over time
- Inform healthcare resource utilisation (HCRU) and costs per health state
- Fill in data gaps for later lines for any comparators which have not been studied in trials (this is not expected to be required).
- Explore the impact of sequencing on effectiveness (this is considered unlikely to be possible).

3.1.2.3. Inclusion and exclusion criteria

The inclusion/exclusion criteria used for identification of RWE are presented in Table 6.

Table 6: Inclusion and exclusion criteria: RWE

PICOS item	Include	Exclude
Population	Studies of participants with advanced (unresectable Stage 3 or Stage 4) RCC	Studies of participants with early stage (not advanced) RCC
Intervention	Any pharmacological treatment for advanced RCC used in the systemic setting	Any pharmacological treatment for advanced RCC not used in the systemic setting
Outcomes	Studies reporting at least one outcome from: • OS • Prognostic variables • PFS • Prognostic variables • Time to progression • TTNT • Time to discontinuation • HRQoL • Current treatment pathways (sequences) being used) • Risk scores • Health costs	Studies not reporting an included outcome
Study design	Real-world evidence	
Other	Geography: UK Time: collection of data within the last 10 years with a focus on datasets including more recent data (2018 onwards)	Geography: Other than UK Time: collection of data > 10 years

Abbreviations: HRQoL, health-related quality of life; OS, overall survival, PFS, progression-free survival; RCC, renal cell carcinoma; RWE, real-world evidence

3.1.2.4. Data extraction and quality assessment strategy

Data extraction of identified RWE was at trial level. Included observational studies were extracted by one reviewer into tables set-up in a word document and checked by a second reviewer. Discrepancies were resolved by discussion, with the involvement of a third reviewer where necessary.

For critical appraisal, ROBINS-I was used to appraise the quality of non-randomised comparative cohort studies. For RWE identified from external datasets, such as patient registries, NICE's Data Suitability Assessment Tool (DataSAT) was completed to provide structured information on data suitability including provenance, quality and relevance.[81] These criteria were considered when conducting quality appraisal.

3.1.2.5. Consultation with clinical experts

As part of its appraisal, the EAG recruited and consulted with three clinical experts in RCC.

- Professor James Larkin, Consultant Medical Oncologist, Royal Marsden Foundation NHS Trust
- Dr Amarnath Challapalli, Consultant Clinical Oncologist, Bristol Cancer Institute, University Hospitals Bristol NHS Foundation Trust
- Dr Teele Kuusk, Urology Consultant, Barts Health NHS Trust

These experts were selected to represent a range of expertise across medical and clinical oncology and urology. The clinical experts were recruited in accordance with the NICE conflict of interest policy.

The following conflicts of interest were declared by the clinical experts:

- Within the last 12 months, Dr James Larkin received honorariums from BMS, Eisai, Merck, Novartis and Pfizer, consultancy fees by BMS, Eisai and Merck, speaker fees from Eisai, Eusa Pharma, Merck, Novartis and Pfizer and institutional research support from BMS, Novartis and Pfizer.
- Within the last 12 months, Dr Amarnath Challapalli received speaker fees and honoraria from Ipsen, BMS, Eisai, Eusa Pharma, Novartis and Pfizer.
- Dr Teele Kuusk declared no conflict of interest for the past 12 months.

To ascertain views on topics such as disease characteristics, typical treatment pathways, disease and treatment outcomes, and treatment effect modifiers (see Appendix M for further details), all three clinical experts took part in a video consultation and provided answers to follow-up questions. Dr Larkin had earlier taken part in another video consultation (prior to the scoping workshop). Expert views were used to provide background information on the condition and on current treatments in UK clinical practice, to guide the methods of this appraisal and to aid interpretation of the appraisal findings.

In addition to this consultation, a broader group of experts (total of 9) were recruited to participate in an expert elicitation exercise to inform long-term OS estimates. This procedure is described in Section 5.2.

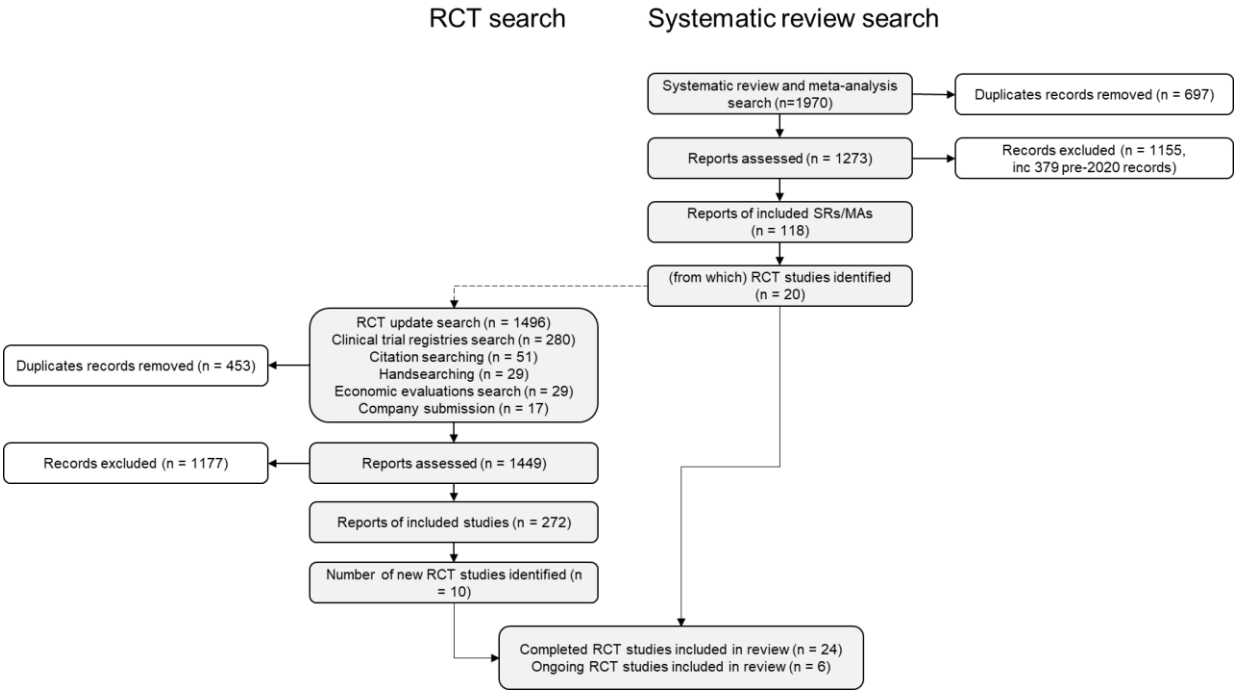
3.1.3. Handling of the company submission

The company submission (CS) was appraised and new information was used to inform the broader project. Specifically, the company's definition of the decision problem and the SLRs and NMAs it conducted were reviewed and compared against the methods used in the EAG's assessment, and references identified by the company were searched to ensure these were identified in the EAG's own searches. New data presented by the company that were not in published reports (for example, new data cuts and information about trial methods contained in the trial clinical study report [CSR]) were extracted and included in our appraisal and analyses. Most prominently, the CS included a new data cut from CheckMate-9ER with data up to a median of 44-months. The company provided Excel files for the relevant time to event endpoints, specifying the number of events and censors per endpoint for PFS, OS, TTD and TTP that were used in the EAG's NMAs and economic model. An appraisal of the company's definition of the decision problem, the methods used in their SLR and analyses, and the latest results from CheckMate 9ER is presented in Section 3.4. A comparison of the company NMA versus the EAG NMA can be found in Section 3.7.7.2.

3.2. Results of the searches for systematic literature reviews and randomised controlled trials

Figure 8 provides an overview of the clinical review searches for SLRs and RCTs. PRISMA diagrams for the individual SLR and RCT searches can be found in Appendix B. In total, 118 SLRs and meta-analyses were identified, and 30 RCTs—20 identified from the SLRs, and a further ten from the RCT top up search and other supplementary search techniques.

Figure 8: Overview of clinical effectiveness searches



Abbreviations: RCT, randomised control trials; SR, systematic reviews; MA, meta-analyses

3.3. Critique of trials identified in the review

3.3.1. Included studies

In total, 30 trials were identified for inclusion in the review. Of these, six are ongoing and are addressed below in Section 3.8. The remaining 24 trials are described below and summarised in Table 7.

Table 7: Clinical evidence included

Study name	Lead reference	Population	Clear cell type (%)	Risk score (IMDC or MSKCC)	Trt line	Comparison
ASPEN (NCT01108445)	Armstrong 2016, Lancet Oncol[84]	Advanced and Metastatic (N=108)	0	Mixed	1L*	suni vs evero
AXIS (NCT00678392)	Rini 2011, Lancet[85]	Advanced and Metastatic (N=723)	100	Mixed	2L	axi vs sora
BERAT (EUDRACT 2011-005939-78)	Grunwald 2022, Oncol Res Treat[86]	Metastatic (N=22)	NR	NR	2L	TKI (axi & suni) vs evero
BIONIKK (NCT02960906)	Vano 2022, Lancet Oncol[87]	Metastatic (N=202)	100	Mixed	1L+	nivo vs nivo+ipi, nivo+ipi vs VEGFR-TKI (suni+pazo)
CABOSUN (NCT01835158)	Choueiri 2018, Eur J Cancer[88]	Metastatic (N=157)	100	Intermediate and poor	1L	cabo vs suni
CheckMate 025 (NCT01668784)	Motzer 2015, NEJM[89]	Advanced and Metastatic (N=821)	100	Mixed	2L and 3L	nivo vs evero
CheckMate 214 (NCT02231749)	Motzer 2018, NEJM[90]	Advanced and Metastatic (N=1096)	100	Mixed	1L	nivo+ipi vs suni
CheckMate 9ER (NCT03141177)	Choueiri 2021a, NEJM[59]	Advanced and Metastatic (N=651)	100	Mixed	1L	cabo+nivo vs suni
CLEAR (NCT02811861)	Motzer 2021b, NEJM[45]	Advanced and Metastatic (N=1069)	100	Mixed	1L	pem+lenv vs lenv+evero vs suni
COMPARZ (NCT00720941)	Motzer 2013, NEJM[91]	Metastatic (N=1110)	100	Mixed	1L	pazo vs suni
CROSS-J-RCC (NCT01481870)	Tomita 2020, Clin Genitourin Cancer[92]	Metastatic (N=120)	100	Favourable and intermediate	1L	suni vs sora
ESPN (NCT01185366)	Tannir 2016, Eur Urol[93]	Metastatic (N=72)	16.7	Mixed	1L*	evero vs suni
Hutson et al, 2017 (NCT00920816)	Hutson 2013, Lancet Oncol[94]	Metastatic (N=288)	100	Favourable and intermediate	1L*	axi vs sora
JAVELIN RENAL 101 (NCT02684006)	Motzer 2019, NEJM[95]	Advanced and Metastatic (N=886)	100	Mixed	1L	ave+axi vs suni

Study name	Lead reference	Population	Clear cell type (%)	Risk score (IMDC or MSKCC)	Trt line	Comparison
METEOR (NCT01865747)	Choueiri 2015, NEJM[96]	Advanced and Metastatic (N=658)	100	Mixed	2L and 3L	cabo vs evero
NCT01136733 (NCT01136733)	Motzer 2015, Lancet Oncol[97]	Advanced and Metastatic (N=153 (101 relevant))	100	Mixed	2L	lenv+evero vs evero
RECORD-1 (NCT00410124))	Motzer 2008 Lancet[98]	Metastatic (N=410)	100	Mixed	2L and 3L	evero vs placebo
RECORD-3 (NCT00903175)	Motzer 2014 J Clin Oncol[99]	Metastatic (N=471)	85	Mixed	1L*	suni vs evero
SWITCH (NCT00732914)	Eichelberg 2015 Eur Urology[100]	Advanced and Metastatic (N=365)	87	Favourable and intermediate	1L	suni vs sora
SWITCH II (NCT01613846)	Retz 2019 Eur J Cancer[101]	Advanced and Metastatic (N=377)	87	Favourable and intermediate	1L	pazo vs sora
SWOG 1500 (NCT02761057)	Pal 2021 Lancet[102]	Advanced and Metastatic (N=152 (94 relevant))	0	Mixed	1L¤	cabo vs suni
TIVO-1 (NCT01030783)	Motzer 2013 J Clin Oncol[103]	Metastatic (N=517)	100	Favourable and intermediate	1L and 2L	tivo vs sora
TIVO-3 (NCT02627963)	Rini 2020 Lancet Oncol[104]	Metastatic (N=350)	100	Mixed	3L and 4L	tivo vs sora
VEG105192 (NCT00334282)	Sternberg 2010 J Clin Oncol[105]	Advanced and Metastatic (N=435)	100	Favourable and intermediate	1L and 2L¥	pazo vs placebo

Abbreviations: RCT, randomised controlled trial; TKI, tyrosine kinase inhibitor; trt, treatment; VEGFR, vascular endothelial growth factor receptors; vs, versus

Notes:

* These trials are not included in the 1st line networks as they do not contain two treatments (or one treatment and a linking treatment) which can be used at 1st line in England and Wales

+ This trial is not currently included in the 1st line network because it includes a non-standard design

¥ This trial is not included in the 1st line network as no other trials compared to placebo and therefore inclusion did not add any value to the network

¤ This trial is not included in the 1st line network as the definition of PFS is not consistent with other trials and given a different histological profile

3.3.2. Description and critique of the design of the studies

Of the 24 included RCTs, the earliest participants were recruited in 2006, with the most recent data cuts in published records drawing from December 2019. Trials included as few as three and as many as 200 centres, with at least 14 trials including UK centres; and had sample sizes across arms comparing relevant treatments of between 22 and 1,110 participants.

Based on an initial consideration of relevant treatments mapped against lines, 18 studies reporting treatments tested at relevant lines were prioritised for inclusion in the review and 8 studies were de-prioritised. Thus, for example, a trial reporting a test at 1st line of a treatment reimbursed only at 2nd line would have been deprioritised. In one situation (NCT01136733), we deprioritised a trial arm in a three-arm trial but retained the relevant comparison.

3.3.2.1. Design of the studies

An overview of study design characteristics for the included trials is shown in Table 8. Of the 24 included trials, 18 were parallel trials and six were crossover trials. The six crossover trials sought to test two-drug sequences characterised by treatment with the first drug to progression; for example, in SWITCH,[100] patients were randomised to sunitinib followed by sorafenib after progression, or sorafenib followed by sunitinib after progression. All 18 parallel trials tested individual treatments to progression or death, with post-progression treatment generally not directly specified, though in six studies[84,98,103,105-107] receipt of the comparator treatment after progression was permitted. In two of these studies (RECORD-1 and VEG105192), this was a crossover from placebo to the comparator treatment.

Though some RCTs included independent masked review (e.g. of progression status), 20 trials were described by study authors as open-label; the remaining trials were distributed as one double-blind, two single-blind, and one triple-blind. Though three trials did not provide sufficient information, 21 trials used stratified randomisation, generally based on risk category and, where relevant, prior treatment.

Only one trial did not report any industry funding (SWOG 1500).

Table 8: Study design characteristics of included trials

Trial name	Line	Comparison	Design (Blinding)	Study sponsor	Continent: Country	Number of centres (number UK centres)	Enrolment period	Final follow-up	Date of last datacut
Prioritised									
1L									
CABOSUN	1L	cabo vs suni	Parallel (Single blind)	Industry and non-industry	North America: USA	77 (0)	Not stated	Median 34.5 months	September 2016
CheckMate 214*	1L	nivo+ipi vs suni	Parallel (Open label)	Industry	Mixed: USA, Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, Columbia, Czechia, Denmark, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Japan, RO Korea, Mexico, Netherlands, Poland, Spain, Sweden, Taiwan, Turkey, UK	175 (6)	Oct 2014 - Feb 2016	67.7 months	February 2021
CheckMate 9ER	1L	cabo+nivo vs suni	Parallel (Single blind)	Industry	Mixed: USA, Europe, Rest of World	125 (3)	Not stated	44 months	May 2022
CLEAR	1L	pem+lenv vs lenv+evero vs suni	Parallel (Open label)	Industry	Mixed: USA, Australia, Austria, Belgium, Canada, Czechia, France, Germany, Greece, Ireland, Israel, Italy, Japan, RO Korea, Netherlands, Poland, Russian Federation, Spain, Switzerland, UK	200 (8)	Oct 2016 - July 2019	49.8 months	August 2020
COMPARZ	1L	pazo vs suni	Parallel (Open label)	Industry	Mixed: North America, Europe, Australia, Asia	Not stated (Not stated)	Aug 2008 - Sept 2011	34.1 months	May 2012
CROSS-J-RCC	1L	suni vs sora	Crossover (Open label)	Industry and non-industry	Asia: Japan	39 (0)	Feb 2010 - July 2012	NR; KM >48 months	June 2015
JAVELIN RENAL 101	1L	ave+axi vs suni	Parallel (Open label)	Industry	Mixed: USA, Canada, Western Europe, Rest of the World	144 (7 investigators, but NR how many centres)	Mar 2016 - Dec 2017	34.1 months	April 2020
SWITCH	1L	suni vs sora	Crossover (Open label)	Industry and	Europe: Germany, Austria, Netherlands	72 (0)	Feb 2009-Dec 2011	15 months	January 2014

Trial name	Line	Comparison	Design (Blinding)	Study sponsor	Continent: Country	Number of centres (number UK centres)	Enrolment period	Final follow-up	Date of last datacut
				non-industry					
SWITCH II	1L	pazo vs sora	Crossover (Open label)	Industry and non-industry	Europe: Germany, Austria, Netherlands	67 (0)	Jun 2012-Sep 2016	NR; KM >45 months	November 2016
TIVO-1*	1L & 2L	tivo vs sora	Parallel (Open label)	Industry	Mixed: USA, Argentina, Bulgaria, Canada, Chile, Czechia, France, Hungary, India, Italy, Poland, Romania, Russian Federation, Serbia, Ukraine, UK	76 (3)	Feb 2010 - Aug 2010	30 months	December 2011
2L +									
AXIS	2L	axi vs sora	Parallel (Open label)	Industry	Mixed: USA, Australia, Austria, Brazil, Canada, China, France, Germany, Greece, India, Ireland, Italy, Japan, RO Korea, Poland, Russian Federation, Singapore, Slovakia, Spain, Sweden, Taiwan, UK	175 (11)	15/09/08 - 23/07/10	37 months	November 2011
BERAT*	2L	TKI (axi/suni) vs evero	Crossover (Open label)	Industry	Europe: Germany	5 (0)	Nov 2012 - Aug 2016	NR' KM curve up to 800 days	January 2020
CheckMate 025*	2L+	nivo vs evero	Parallel (Open label)	Industry	Mixed: USA, Argentina, Australia, Austria, Belgium, Brazil, Canada, Czechia, Denmark, Finland, France, Germany, Greece, Ireland, Israel, Italy, Japan, Norway, Poland, Romania, Russian Federation, Spain, Sweden, UK	146 (5)	Oct 2012 - Mar 2014	72 months	NR
METEOR	2L+	cabo vs evero	Parallel (Open label)	Industry	Mixed: Multiple	173 (11)	Aug 2013 - Nov 2014	18.8 months	December 2015
NCT01136733	2L+	lenv+evero vs evero	Parallel (Open label)	Industry	Mixed: Czech Republic, Poland, Spain, UK, USA	37 (11)	March 2012-June 2013	approx. 24 months median at	December 2014

Trial name	Line	Comparison	Design (Blinding)	Study sponsor	Continent: Country	Number of centres (number UK centres)	Enrolment period	Final follow-up	Date of last datacut
								follow-up	
RECORD-1*	2L+	evero vs placebo	Parallel (Double blind)	Industry	Mixed: Australia, Canada, Europe, Japan, USA	86 (NR)	Nov 2006-Nov 2007	21 months	November 2008
TIVO-3	3L+	tivo vs sora	Parallel (Open label)	Industry	Mixed: USA, Belgium, Canada, Czechia, Denmark, France, Germany, Hungary, Italy, Poland, Spain, UK	120 (17)	May 2016 - Aug 2017	NR; KM up to 48 months	May 2021
Deprioritised									
ASPEN	1L	suni vs evero	Parallel (Open label)*	Industry and non-industry	Mixed: USA, Canada, UK	17 (6)	23/09/2010 - 28/10/2013	29 months	May 2016
BIONIKK	1L	nivo vs nivo+ipi, nivo+ipi vs VEGFR-TKI (suni/pazo)	Parallel (Open label)	Industry	Europe: France	15 (0)	28/06/2017 - 18/07/2019	Median 42.1 months (40.5 - 45.2)	NR
ESPN*	1L	evero vs suni	Crossover (Open label)	Industry and non-industry	North America: USA	3 (0)	Not stated	23.6 months	May 2014
Hutson et al, 2017	1L	axi vs sora	Parallel (Open label)	Industry	Mixed: USA, Mexico, Asia, Eastern Europe	126 (0)	Jun 2010 - Apr 2011	4.5 years	December 2014
RECORD-3*	1L	suni vs evero	Crossover (Open label)	Industry	Mixed: USA, Argentina, Australia, Brazil, Canada, Denmark, France, Germany, Hong Kong, Italy, South Korea, Mexico, Netherlands, Peru, Spain, Taiwan, Thailand, Turkey, United Kingdom	83 (3)	Oct 2009-Jun 2011	Median 3.7 years	May 2015
SWOG 1500	1L*	cabo vs suni	Parallel (Open label)	Non-industry	North America: USA, Canada	65 (0)	April 2016-Dec 2019	NR; KM to 40 months	October 2020

Trial name	Line	Comparison	Design (Blinding)	Study sponsor	Continent: Country	Number of centres (number UK centres)	Enrolment period	Final follow-up	Date of last datacut
SUNNIFORECAST	1L	nivo+ipi vs SoC	Parallel (Open label)	Industry	Europe: Belgium, Czech Republic, France, Germany, Netherlands, Spain, United Kingdom	30 (2)	Nov 2017-ongoing	NR	NR
VEG105192*	1L and 2L	pazo vs placebo	Parallel (Triple blind)	Industry and non-industry	Mixed: Argentina, Australia, Austria, Brazil, Chile, China, Czech Republic, Estonia, Greece, Hong Kong, India, Ireland, Italy, RO Korea, Latvia, Lithuania, Mexico, New Zealand, Pakistan, Poland, Russian Federation, Slovakia, Tunisia, Ukraine, UK	80 (5)	Apr 2006-Apr 2007	Unclear	March 2010

Abbreviations: NR, not reported; SD, standard deviation; UK, United Kingdom; USA, United States of America; vs, versus

*Crossover to the comparator permitted following progression

3.3.2.2. Population

Inclusion/ exclusion criteria

Included trials included participants aged 18 years or older with histologically confirmed RCC, measurable via RECIST guidelines, and with participants having adequate performance status (generally defined as ECOG performance status of 0 or 1, or as Karnofsky Performance Score of 70% or above). All trials required participants to have advanced or metastatic RCC, though the exact form of wording varied including within different reports of the same trial. Exclusion criteria related principally to other health parameters, such as controlled hypertension and adequate organ function; in addition, most trials reported explicit exclusion criteria with respect to brain and central nervous system metastases.

Additional criteria related principally to prior lines of treatment and risk group. These are discussed under baseline characteristics.

Baseline characteristics

An overview of the sample characteristics in the prioritised and deprioritised trials is shown in Table 9.

Histology. Of the 24 trials, 17 included patients with clear cell RCC only, or RCC with a clear cell component. Studies with a whole or majority (>85%) clear cell component were prioritised for inclusion. Three trials that were prioritised and two that were de-prioritised included participants with both clear cell and non-clear cell RCC.[86,93,99-101] The remaining three trials specifically targeted participants with predominantly non-clear-cell RCC histology.[84,108]

Risk distribution. Risk distribution was measured by a combination of IMDC and MSKCC risk scores. For convenience, both sets of risk scoring methods are described as producing risk score classes as ‘favourable’, ‘intermediate’ or ‘poor’. Two prioritised trials[86,92] did not enrol any participants assessed as having poor risk, and a further three prioritised[100,101,103] and two de-prioritised trials[94,105] enrolled a very low number of participants assessed as being at poor risk (i.e. ≤5% of the trial sample). One prioritised trial[88] only enrolled participants assessed as being at intermediate or poor risk. Proportions of participants assessed as being at favourable risk ranged in trials from 0 to 52%, while for intermediate risk, participants proportions ranged from 37% to 81%. Proportions of participants assessed as being at poor risk ranged from 0% to 40%.

Prior lines of systemic therapy. Of 24 trials, 17 RCTs included participants for whom the study drug was classed as their 1st line of systemic therapy. Of these 17 trials, 14 were only in participants receiving 1st line treatment. The remaining three trials enrolled patients to receive 1st line and 2nd line treatments; for these trials, the proportion of patients receiving their first systemic treatment ranged from 93% to 53%. Ten trials in the 1st line setting were prioritised for inclusion.

Correspondingly, 10 trials enrolled participants receiving 2nd line or later therapy. Distinguishing between participants receiving 2nd line and 3rd line systemic treatments was complicated by the fact that trials inconsistently included participants on the basis of prior lines of treatment belonging to a specific class. However, data presented in included studies indicated that beyond three trials enrolling a mix of 1st line and 2nd line patients, an additional two trials enrolled only participants for the 2nd line of treatment. Of the remaining five trials, four enrolled participants across 2nd line and 3rd line, with ranges of 2nd line treatment between 20% and 72%; and one trial enrolled only participants at the 3rd and 4th lines of therapy, with 60% of participants at 3rd line. Seven trials in the 2nd line-plus setting were prioritised for inclusion.

Prior systemic TKI or immunotherapy. Data on the proportions of participants with prior systemic TKI were inconsistently reported. All of the 11 trials that reported data on prior TKI use were prioritised for inclusion, and included five trials[96-98,104,107] that enrolled only participants with prior TKI, five trials[88,91,106,109,110] that enrolled participants only without prior TKIs, and one trial[85] that enrolled a blend of participants with and without prior TKI. Data on the proportions of participants with prior immunotherapies were also inconsistently reported. All of the 12 trials reporting data on this point were prioritised for inclusion, and included six trials with no participants who had previously received prior immunotherapies.

Prior surgery. Data on prior nephrectomy were reported for 22 trials, of which 17 were prioritised for inclusion. One prioritised trial[103] enrolled only participants with prior nephrectomy. In two trials[86,93] (one prioritised and one deprioritised), a minority of participants had previously undergone nephrectomy. In all other trials, the vast majority of participants (more than two thirds) had undergone nephrectomy prior to the trial.

Table 9: Population characteristics of included trials

Trial name	N (UK pts)	Key inclusion and exclusion criteria						Baseline characteristics					
		Age	Hist	Risk	Prior trt	ECOG	Other	Median age (range) years	Histology (% clear cell; % sarco features)	≥ 2 met sites	% bone mets	% risk status: Fav; Int; Poor	% prior nephrectomy
Prioritised													
1L													
CABOSUN	157 (NR)	≥18	CC	I/P	None	0-2	Pts with known brain mets: adequately treated and stable for 3 months	63.0 (31, 87)	100 / NR	72.6	36.3	0; 81; 19	74.5
CheckMate 214	1096 (NR)	≥18	CC	-	None	KPS≥70	Exclusion: CNS mets or auto immune disease & glucocorticoid or immunosuppressant use	62 (21 - 85)	100 / 13	78	21.1	23; 61; 16	81.2
CheckMate 9ER	651 (21)	≥18	CC	-	None	KPS≥70	One previous adjuvant or neoadjuvant therapy Exclusion: active CNS, active autoimmune disease	Cabo+nivo 62 (29-90). Suni 61 (28-86)	100 / 11.9	71.7	23.0	23; 57; 20	69.9
CLEAR	1069 (NR)	≥18	CC	-	None	KPS≥70	Exclusion: unstable CNS mets, active autoimmune disease in the past 2 years	Pem+lenv 64 (34–88), lenv+evero 62 (32–86), suni 61 (29–82)	100 / 6.8	68.8	25.1	32; 55; 10	74.6
COMPARZ	1110 (NR)	≥18	CC	-	None	KPS≥70	Exclusion: brain mets, poorly controlled hypertension	Pazo 61 (18-88), suni 62 (23-86)	100 / NR	38.3	17.6	27; 59; 11	83.2
CROSS-J-RCC	120 (0)	18-80	-	F/I	None	0-2	Exclusion: unstable brain mets (not stable 2 months before screening)	67 (41-79); suni first 67 (41-79), sora first 66 (44-79)	100 / NR	92.5	28.3	21.7; 78.3; 0	88.3
JAVELIN RENAL 101	886 (NR)	≥18	CC	-	None	0 or 1	Exclusion: active CNS mets, autoimmune disease, and current or previous use of	Ave+axi 62.0 (29.0-83.0). suni 61.0 (27.0 -88.0)	100 / 12	58.2	23.3	22; 65; 11	81.7

Trial name	N (UK pts)	Key inclusion and exclusion criteria						Baseline characteristics					
		Age	Hist	Risk	Prior trt	ECOG	Other	Median age (range) years	Histology (% clear cell; % sarco features)	≥ 2 met sites	% bone mets	% risk status: Fav; Int; Poor	% prior nephrectomy
							glucocorticoid or immunosuppressants 7 days before randomization						
SWITCH	365 (0)	18-85	-	F/I	None	0 or 1	Unsuitable for cytokine therapy Exclusion: symptomatic met brain tumours	65 (39-84)	87 / NR	64	15	42; 55; 0.5	92
SWITCH II	377 (0)	18-85	-	F/I	None	KPS≥70	Unsuitable for cytokine therapy Exclusion: uncontrolled brain mets	68 (26-86)	87 / NR	NR	20	49; 48; 2	99
TIVO-1	517 (4)	≥18	CC	-	0 or 1	0 or 1	Prior nephrectomy Exclusion: prior VEGF Unstable brain mets ≥ 3 months following prior treatment	59 (23 - 85)	100 / NR	68.3	21.9	30; 65; 5	100
2nd line +													
AXIS	723 (NR)	≥18	CC	-	1*	0 or 1	Life expectancy of ≥12 weeks Exclusion: CNS mets	NR for whole sample	100 / NR	NR	NR	20; 64; 10	91
BERAT	22 (0)	NR	-	F/I	NR	0 or 1	CNS mets were permitted if local treatment was completed ≥3 months, and steroids were discontinued	55.3	NR / NR	90	10	NR; NR; 0	20
CheckMate 025	821 (26)	≥18	CC	-	1-2	KPS≥70	Exclusion: CNS mets Condition treated with glucocorticoids (equivalent to >10 mg of prednisone daily)	62 (18–88)	100 / NR	83	18	36; 49; 15	88

Trial name	N (UK pts)	Key inclusion and exclusion criteria						Baseline characteristics					
		Age	Hist	Risk	Prior trt	ECOG	Other	Median age (range) years	Histology (% clear cell; % sarco features)	≥ 2 met sites	% bone mets	% risk status: Fav; Int; Poor	% prior nephrectomy
METEOR	658 (26)	≥18	CC	-	≥1 TKI	KPS≥70	Disease progression during or within six months of the most recent VEGFR/TKI treatment, and within 6 months before randomisation Pts with known brain mets that were adequately treated and stable were eligible	Cabo 63 (32-86), evero 62 (31-84)	100 / NR	81.5	22	46; 42; 13	85
NCT01136733	101 (50)	≥18	CC	-	1 TKI	0 or 1	Within 9 months of stopping previous treatment Exclusion: brain mets	61, 37-79	100 / NR	79	07	23; 37; 40	88
RECORD-1	410 (NR)	≥18	CC	-	≥1	KPS≥70	Progressed on or within 6 months of stopping treatment with suni or sora, or both drugs Previous therapy with bev, IL2, or IFNα permitted Exclusion: untreated CNS mets	61, 27-85	100 / NR	91	35	29; 56; 14	97
TIVO-3	350 (NR)	≥18	CC	-	2 or 3¥	0 or 1	Life expectancy ≥3 months Exclusion: CNS mets (other than lesions that were radiographically stable without any steroid treatment for at ≥3 months)	63 (30, 90)	100 / NR	89.1	NR	21; 61; 18	NR

Trial name	N (UK pts)	Key inclusion and exclusion criteria						Baseline characteristics					
		Age	Hist	Risk	Prior trt	ECOG	Other	Median age (range) years	Histology (% clear cell; % sarco features)	≥ 2 met sites	% bone mets	% risk status: Fav; Int; Poor	% prior nephrectomy
Deprioritised trials													
ASPEN	108 (NR)	≥18	nCC	-	None	KPS≥60	Life expectancy ≥3 months Exclusion: active untreated CNS mets	63 (23, 100)	0 / 14.8	NR	25	27; 60; 14	79.6
BIONIKK	202 (0)	≥18	NR	-	None	0-2	Exclusion: uncontrolled or symptomatic brain mets	Medians across groups ranged from 59 to 66	100 / 26.6	74.4	20.6	30; 50; 20	NR
ESPN	72 (0)	≥18	Mix¤	-	None	0 or 1	Exclusion: untreated brain metastases	Evero 58 (23–73), suni 60 (28–76)	16.7 / 26	82.4	26	10; 74; 16	47.1
Hutson et al, 2017	288 (0)	≥18	CC	-	None	0 or 1	Life expectancy 12 weeks Exclusion: brain mets or CNS involvement	Axi 58·0 (23–83), sora 58·0 (20–77)	100 / NR	NR	27.8	51; 43; 3	86.8
RECORD-3	471 (NR)	≥18	Mix	-	None	KPS≥70	Exclusion: CNS mets	62 (20-89)	85 / NR	68	23	29; 56; 15	67
SWOG 1500	90 (0)	≥18	nCC	-	0 o 1	Zubrod PS 0 - 1	Patients with known brain mets who had received adequate treatment were eligible Exclusion: prior treatment with excluding VEGF-directed or MET-directed drugs	65 (58-75)	Papillary RCC 0 / NR	NR	14.4	26; 61; 14	73.3
SUNNIFO RECAST	237 (NR)	≥18	nCC	-	None	KPS≥70	Exclusion: ccRCC component >50% Active brain mets requiring systemic corticosteroids	NR for whole sample	148 papillary, 83 non-papillary, 0 clear cell; sarcomatoi	NR	NR	NR; NR; NR	NR

Trial name	N (UK pts)	Key inclusion and exclusion criteria						Baseline characteristics					
		Age	Hist	Risk	Prior trt	ECOG	Other	Median age (range) years	Histology (% clear cell; % sarco features)	≥ 2 met sites	% bone mets	% risk status: Fav; Int; Poor	% prior nephrectomy
									d features NR				
VEG105192	435 (NR)	≥18	CC	-	0 or 1 +	0 or 1	Exclusion: CNS mets	NR for whole sample	100 / NR	83.2	27.4	39; 54; 3	88.5

Abbreviations: ave, avelumab; axi, axitinib; bev, bevacizumab; cabo, cabozantinib; CC, clear cell; CNS, central nervous system; ECOG, Eastern Cooperative Oncology Group; evero, everolimus; fav, favourable; IFNα, interferon alpha; IL-2, interleukin 2; int, intermediate; KPS, Karnofsky Performance Status; lenv, lenvatinib; MET, mesenchymal-epithelial transition; mets, metastasis; NA not applicable; nCC, non clear cell; nivo, nivolumab; NR not reported; pazo, pazopanib; pem, pembrolizumab; PS, performance status; Pts, patients; RCC, renal cell carcinoma; sora, sorafenib; suni, sunitinib; TKI, tyrosine kinase inhibitors; trt, treatment; UK, United Kingdom; VEGF, vascular endothelial growth factor

* RECIST-defined progressive disease as assessed by investigators after one previous systemic 1L regimen with a suni-based, bevacizumab + interferon-alfa-based, temsirolimus-based, or cytokine-based regimen, 2 weeks or more since end of previous systemic treatment (4 weeks or more for bevacizumab + interferon-alfa),

¥ one of which included a VEGFR TKI other than tivo or sora

¤ Advanced papillary, chromophobe, collecting duct carcinoma, Xp11.2 translocation, unclassified RCC, or ccRCC with >20% sarcomatoid features in their primary tumours

+ progressed on one prior cytokine-based systemic therapy (amended to include treatment-naive patients living in countries where there were barriers to the access of established therapies or where cytokines were not recognized as standard treatment for RCC)

3.3.2.3. Interventions and comparators

An overview of the intervention characteristics used in the included trials is shown in Table 10. Interventions and comparators were distributed unevenly across the included trials. Our commentary focuses here only on relevant arms in included trials. There was evidence from at least one trial for all relevant active interventions. No trials used 'current care', investigator's choice or best supportive care as a comparator, but placebo was used as a comparator in two trials,[98,105] one of which was prioritised for inclusion. Sunitinib was the most commonly represented treatment. An overview of interventions is as follows:

- Sunitinib: 14 trials (10 prioritised)
- Single-agent everolimus: 8 trials (5 prioritised)
- Sorafenib (used as a linking treatment): 7 trials (6 prioritised)
- Pazopanib: 4 trials (2 prioritised)
- Single-agent axitinib: 3 trials (2 prioritised)
- Single-agent cabozantinib: 4 trials (3 prioritised).
- Single-agent nivolumab: 2 trials (1 prioritised)
- Nivolumab + ipilimumab: 2 trials (1 prioritised)
- Single-agent tivozanib: 2 trials (2 prioritised)
- Lenvatinib + everolimus: 2 trials (2 prioritised)
- Avelumab + axitinib: 1 trial (1 prioritised)
- Cabozantinib + nivolumab: 1 trial (1 prioritised)
- Pembrolizumab + lenvatinib: 1 trial (1 prioritised)

Table 10: Intervention characteristics of included trials

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery etc)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
Prioritised						
1L						
CABOSUN	NA	Cabo vs suni	Cabo (orally): 60mg OD. Suni (orally): 50mg OD for 4 wks then 2-wk break per cycle.	NR	N/A	Int 60.8 Control 61.5
CheckMate 214	NA	Nivo+ipi vs suni	Nivo (IV): 3 mg/ kg bodyweight over 60-minute period/ 3 wks for four doses and then at a dose of 3 mg/ kg bodyweight every 2 wks. Ipi (IV): 1 mg/ kg bodyweight over a period of 30 minutes/ 3 wks for four doses. Suni (orally): 50 mg OD for 4 wks, 2 wks off per cycle. Nivo or ipi dose reductions not allowed. Dose delays for adverse events were permitted in both groups.	Nivo induction: 79*; Nivo maintenance: [redacted] Ipi: 79*	Treated beyond progression: Nivo+ipi n=157 (29%), Suni n=129 (24%)	Int 53.5 Control 66.5
CheckMate 9ER	NA	Cabo+nivo vs suni	Nivo (IV): 240 mg every 2 wks and cabo(orally) 40 mg OD. Suni (orally): 50 mg OD for 4 wks then 2-wk break in 6-wk cycle.	Nivo: [redacted] Cabo: [redacted] Suni: NR	Nivo stopped after 2 years (from the first dose)	Int 25.1 Control 40.5
CLEAR	NA	Pem+lenv vs lenv+evero vs suni	Pem+lenv: for 21-day cycle, lenv (orally) 20 mg OD and pem (IV) 200 mg on day 1 of cycle. Suni (orally): 50 mg OD (4 wks on/2 wks off). Dose reduction and interruptions: investigators decide the probability of the event being related to 1 or both drugs. lenv dose reduction to 14, 10, and 8	Median Pem+lenv Len: 69.6% Median number of pem infusions per patient 22 (range, 1 to 39). Suni 83.2%	Maximum 35 treatments for pem All patients could continue treatment beyond progression if they received clinical benefit and tolerated the study-drug treatment	Int pem+lenv = 32.96 Control 57.7

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery etc)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
			mg/day). Dose reductions below 8 mg/day must be discussed with sponsor.			
COMPARZ	NA	Pazo vs suni	Pazo was administered orally at a once-daily dose of 800 mg, with continuous dosing. Suni was administered orally in 6-week cycles at a once-daily dose of 50 mg for 4 weeks, followed by 2 weeks without treatment. Dose reductions for pazo (to 600 mg and then to 400 mg) and suni (to 37.5 mg and then to 25 mg) were permitted due to adverse events.	NR	N/A	Int NR Control NR
CROSS-J-RCC	NA	Suni vs sora	Suni (orally): 50 mg OD (4 wks on/2 wks off). Suni dose reductions to 37.5mg then 25 mg/day schedule 4/2. Dose reduction below 25 mg/day discussed with the sponsor.	Median RDI - suni 65.8% (range 7.1%-100%), sora 61.2% (range 10.7%-100%)	N/A	Int NR Control NR
JAVELIN RENAL 101	NA	Ave+axi vs suni	Ave+axi: ave (IV) 10mg/kg every 2 wks and axi (orally) 5 mg BID. Suni (orally): 50 mg OD (4 wks on/2 wks off).	Ave 91.5%; Axi 89.4%; Sun 83.9% (all median)	N/A	Int 46.2 Control 60.6
SWITCH	NA	Suni vs sora	Suni (orally): 50 mg OD, 4 wks on 2 wks off; dose reductions permitted	NR	N/A	Int 57% crossed over Control 42% crossed over
SWITCH II	NA	Pazo vs sora	Pazo (orally) 800mg OD, dose reductions permitted	NR	N/A	Int 64.0 Control 58.5
TIVO-1	NA	Tivo vs sora	Tivo (orally) 1.5mg OD for 3 wks followed by 1 wk off per cycle. Specific guidelines for hypertension, otherwise AEs ≥ grade 3 were managed by a dose reduction to 1.0 mg per day.	Tivo 94%; sora 80%	N/A	Int 18.1 Control 64.2
2L+						
AXIS	TKI 54%; IO Cytokines 35%; Bev 8%	Axi vs sora	Axi (orally): 5 mg BID with continuous dosing, if tolerated (no adverse reactions above grade 2 for at least 2 weeks) dose increased to 7 mg twice daily unless the patient's blood pressure was higher than	Median 99% for axi and 92% for sora	Patients were treated until progression of disease (RECIST version 1·017),	Int 54.4 Control 56.6

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery etc)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
			150/90 mm Hg or the patient was receiving anti hypertensive medication. If tolerated, increased to a maximum of 10 mg twice daily. Dose could be reduced to 3 mg twice daily and then further to 2 mg twice daily.		occurrence of unacceptable toxic effects, death, or withdrawal of patient consent	
BERAT	TKI NR; IO NR	TKI (axi/suni) vs evero	Axi: 5mg BID starting dose Suni: 50 mg OD, 4-2 regimen. Evero: 10mg OD	NR	Trial stopped due to poor accrual	Int TKI 60% Control evero 80%
CheckMate 025	TKI 100%; IO NR	Nivo vs evero	Nivo (IV): 3 mg/ kg of body weight as a 60-minute every 2 wks. Evero (orally):10 mg OD. Dose modifications were not permitted for nivo but were permitted for evero.	NR	Continuation after initial disease progression was allowed if the investigator noted that there was a clinical benefit and the study drug had an acceptable side-effect profile	Int 67.3 Control 72.0
METEOR	TKI 100%; IO >7%	Cabo vs evero	Cabo (orally): OD at 60 mg. Evero (orally): OD at 10 mg.	Cabo: NS; Evero 84%	Patients were allowed to continue study treatment beyond radiographic progression at the discretion of the investigator.	Int 50 Control 55
NCT01136733	TKI 100%; IO 3%	Lenv+evero vs evero	Lenv+evero: lenv (18 mg/day) as one 10 mg capsule and two 4 mg capsules + eve (5 mg/day) as one 5 mg tablet. Single-agent evero (10 mg/day) two 5 mg tablets	NR	N/A	Int 27.5 Control 36
RECORD-1	TKI 100%; IO 65%	Evero vs placebo	Evero (orally): 10 mg/d + BSC. Matching placebo plus BSC	NR	N/A	Int NR Control 79.9
TIVO-3	TKI 100%; IO/TKI tivo	Tivo vs sora	Tivo (orally): 1·5 mg OD in 4-wk cycles comprising 21 days on treatment followed by 7 days off treatment. Dose reduction to	NR	N/A	Int 64.6 Control 58.5

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery etc)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
	27%, sora 25%		1.0 mg OD allowed for patients with treatment-related AEs ≥ Grade 3. Dose interruptions allowed for persistent AEs.			
De-prioritised						
ASPEN	TKI NA; IO NA	Suni vs evero	Suni (orally): 50 mg OD on days 1-28 of each 42-day cycle. Dose reductions permitted or recommended for Grade 3 toxic effects and required for Grade 4 toxic effects: reduction to 37·5 mg or 25 mg; holds such as alternative dosing treatment cycles of 2 weeks on treatment and 1 week off treatment, depending on the timing and severity of toxic effects. Evero (orally): 10 mg OD on days 1-42 for each 42 day cycle. Dose reductions permitted or recommended for Grade 3 toxic effects and required for Grade 4 toxic effects: reduction to 5 mg once daily and then to 5 mg every other day.	NR	N/A	Int 71 Control 58
BIONIKK	TKI NR; IO NR	Nivo vs nivo+ipi, nivo+ipi vs VEGFR-TKI (suni/pazo)	Nivo+ipi (IV): nivo 3 mg/kg plus ipi 1 mg/kg every 3 wks for 4 doses then IV nivo 240 mg every 2 wks. Nivo (IV): 240 mg every 2 wks. Suni (orally) 50 OD for 4 wks every 6 wks Pazo (orally) 800 mg OD continuously.	NR	NR	Nivo: 62 Nivo+Ipi: 57.4 TKI: 50
ESPN	TKI NA; IO NA	Evero vs suni	Evero 10 mg/d orally 4 wk on and 2 wk off; suni 50 mg/d orally 4 wk on and 2 wk off	NR	N/A	Int NR Control NR
Hutson et al, 2017	TKI 0; IO 0	Axi vs sora	AXI (orally): 5 mg BID with food, in 4-wk cycles. Doses can be increased first to 7 mg BID, and subsequently to 10 mg BID for patients who had not had any grade 2+ TRAEs for at least 2 wks and had blood pressure ≤150/90 mm Hg. Those with AEs or lab abnormalities could have dose reduced to 3 mg BID, and then 2 mg BID. PD patients who had clinical benefit could continue on treatment	Axi 125%, Sora 98%	NR	Int 15.1 Control 19.8

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery etc)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
RECORD-3	TKI 0; IO 0	Suni vs evero	Evero: 10 mg/day Suni: 50 mg/ day (4 wks on, 2 wks off)	Evero 98%, suni 87%	N/A	Int 55 Control 51
SWOG 1500	NA	Cabo vs suni	Cabo(orally): 60 mg OD, dose reductions permitted, Suni (orally) 50 mg 4 wks on, 2 wks off, dose reductions permitted	NR	N/A	Int NR Control NR
SUNNIFORECAST	TKI 0; IO 0	Nivo+ipi vs SoC	Nivo+ipi: nivo (IV) 3 mg/kg + ipi (IV) 1 mg/kg every 3 wks for 4 doses followed by nivo fixed dose 240 mg IV every 2 wks or fixed dose 480 mg IV every 4 wks	NR	N/A	Int NR Control NR
VEG105192	TKI 0; IO 0	Pazo vs placebo	Pazo (Orally): 800 mg OD Administered 1 hour before or 2 hours after meals. Dose modification guidelines for AEs were prespecified (details not reported).	NR	N/A	Int 30.3 Control 65.5

Abbreviations: AEs, adverse events; ave, avelumab; axi, axitinib; BID, twice daily; BSC, best supportive care; cabo, cabozantinib; evero, everolimus; Int, intermediate; IO, immuno-oncology; ipi, ipilimumab; ITT, intention to treat; IV, intravenous; lenv, lenvatinib; N/A, not applicable; nivo, nivolumab; NR, not reported; OD, once daily; pazo, pazopanib; PD, progressed disease; pem, pembrolizumab; RDI, relative dosing intensity; RECIST, Response Evaluation Criteria in Solid Tumors; SoC, standard of care; sora, sorafenib; suni, sunitinib; tivo, tivozanib; TKI, tyrosine kinase inhibitor; Tx, treatment; UK, United Kingdom; VEGR, vascular endothelial growth factor receptor; vs, versus; wks, weeks

Notes: dosing is only included for treatments which are part of the UK treatment pathway

*79% reported to receive all 4 doses of nivo and ipi within the induction phase

3.3.2.4. Outcomes

The outcomes reported in the 24 trials are summarised in Table 11. The account of outcomes is derived from publicly available trial reports.

Overall survival

OS was measured in all included trials. Details of follow-up duration were reported for 17 trials, and in a range of ways. Where trials reported the time to final follow-up (n=8), this was below two years in one case and up to seven years in one case; five trials had final follow-ups of between two and four years. An additional trial reported minimum follow-up of 13 months. The remaining eight trials reported median or average follow-up period. Four trials reported median or average follow-up of less than two years, one a median follow-up of two years and a final three trials a median follow-up of between three and six years. Because most analysis protocols were event-driven and included interim analyses, OS data were of variable maturity between trials, highlighting the need for extrapolation.

Adjustment for crossover and treatment-switching was inconsistently addressed in included trials. In trials with a crossover design, OS was not adjusted as the goal of the analysis was to capture the crossover between two different drugs. Treatment-switching adjustments to OS were reported in relatively few trials. Where subsequent treatments were reported, these were inconsistently aligned with UK practice, often making use of treatments (e.g. sorafenib) that are not part of UK treatment pathways. Information on subsequent treatments forming sequences that would be 'disallowed' in UK practice (e.g. immuno-oncology therapies followed by immuno-oncology therapies) was only inconsistently presented across trials.

Progression-free survival

PFS on first treatment was also included in all 24 trials. 23 of 24 trials used a standard definition for PFS of time to the first of RECIST-assessed progression or death. One trial (SWOG 1500) used a non-standard definition which included clinical progression and symptomatic deterioration (investigator assessed). Where PFS censoring rules were mentioned in trial protocols the trials specified FDA analysis rules where patients are censored on receipt of subsequent treatment if this is prior to progression. It is noted that that EAG in TA858 performed sensitivity analysis looking at the use of EMA rules which count receipt of subsequent treatment as an event. These are analyses are redacted and the amount of difference this made to the appraisal is unclear. 10 trials assessed PFS via blinded independent central review (BICR), 2

used an independent review committee with no or unclear blinding and the remaining 12 were investigator-assessed. All of the combination therapy trials were assessed via independent central review except CheckMate 214.

Because an important element of PFS is monitoring of disease status, the tumour scan frequency used in the trials were extracted. In the 20 trials reporting tumour scan frequency, seven used a based frequency of eight weeks, and six used a base frequency of every 12 weeks or three months (with one including an interim scan after six weeks on treatment). Two trials scanned every eight weeks in the first year of study treatment with every 12 weeks thereafter. Two trials scanned 12 weeks after randomisation, then took scans every six weeks for a period of time (up to 13–14 months post-randomisation) and then every 12 weeks thereafter. Two trials scanned at Weeks 6 and 12, and then every eight weeks. One trial scanned every six weeks until Week 12 and then every eight weeks until progression. Three trials described additional scan frequency related to bone and brain metastases where relevant.

Additional time-to-event outcomes

Four trials reported TTP outcomes in publicly available trial reports, including one reporting time to deterioration on treatment as a composite outcome. Three trials also reported time to next treatment outcomes. Six trials reported time to discontinuation.

Duration of response and response rate

Duration of response was reported in 13 trials. Response rate was reported in 24 trials.

Adverse events

The incidence and prevalence of AEs were reported in some form for all 24 trials. This generally included reporting of most common adverse events, though discontinuation due to AEs was also reported for nearly all trials in some form.

Health-related quality of life

HRQoL outcomes were identified for 16 trials. Utility data identified are presented in the later sections relevant to the economic analysis (Section 4.3.7.1).

Table 11: Outcomes reported by RCTs included in the review

Trial name	OS	PFS	TTP	TTNT	TTD	Duration of response	Response rate	Adverse events	HRQoL
ASPEN	X	X					X	X	X
AXIS	X	X				X	X	X	X
BERAT	X	X					X	X	X
BIONIKK	X	X		X		X	X	X	
CABOSUN	X	X			X		X	X	
CheckMate 025	X	X	X			X	X	X	X
CheckMate 214	X	X		X	X	X	X	X	X
CheckMate 9ER	X	X		X	X	X	X	X	X
CLEAR	X	X				X	X	X	X
COMPARZ	X	X					X	X	X
CROSS-J-RCC	X	X			X	X	X	X	
ESPN	X	X					X	X	
Hutson et al, 2017	X	X	X*			X	X	X	X
JAVELIN RENAL 101	X	X				X	X	X	¥
METEOR	X	X					X	X	X
NCT01136733	X	X				X	X	X	
RECORD-1	X	X					X	X	X
RECORD-3	X	X				X	X	X	X
SWITCH	X	X	X		X		X	X	
SWITCH II	X	X	X		X		X	X	X
SWOG 1500	X	X					X	X	
TIVO-1	X	X					X	X	X
TIVO-3	X	X				X	X	X	
VEG105192	X	X				X	X	X	X
TOTAL	24	24	4	3	6	13	24	24	16

Abbreviations: HRQoL, health-related quality of life; OS, overall survival; PFS, progression free survival; RCT, randomised controlled trial; TTD, time to discontinuation; TTNT, time to next treatment; TTP, time to progression.
Notes: *Time to treatment failure ¥ utility data reported within the economics section of TA645 but not clinical outcomes reported by arm

3.3.2.5. Critical appraisal of the included studies

The full quality assessments of RCTs included in this appraisal are presented in Table 12. A summary of bias issues across the trials is provided in the following section. None of the included trials were appraised as being at a low overall risk of bias. Of the seventeen prioritised trials, five were appraised as being at a high risk of bias and twelve were appraised as being at an unclear risk of bias.

Table 12: Summary of domain-level risk of bias judgments, main issues per study and overall study-level risk of bias

Trial name	Overall line	Selection bias	Performance and detection biases	Attrition bias	Reporting bias	Conflict of interest	Other bias	Overall study-level risk of bias	Main issues
Prioritised									
AXIS	2L	Low	High	Unclear	Low	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, potential conflict from industry funding
BERAT	2L	Unclear	High	High	Unclear	High	High	**High**	Unclear reporting of randomisation and allocation concealment, small sample with potential baseline imbalances, open-label trial with some highly subjective outcomes, very high differential attrition with no methods to account for missing data, the paper reported on more outcomes than were listed in the trial registry, potential conflict from industry funding, risk of carryover effect as no washout period is specified
CABOSUN	1L	High	Unclear	Unclear	Low	High	Low	**High**	Dynamic allocation of treatment, open-label trial with some subjective outcomes, very high but non-differential attrition with inadequate methods to account for missing data, potential conflict from industry funding
CheckMate 025	2L and 3L	Low	High	Unclear	Low	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, some imbalances in attrition by reason with inadequate methods to account for missing data, potential conflict from industry funding
CheckMate 214	1L	Low	High	Unclear	Low	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, some imbalances in attrition by reason with methods to account for missing data unclear, potential conflict from industry funding

Trial name	Overall line	Selection bias	Performance and detection biases	Attrition bias	Reporting bias	Conflict of interest	Other bias	Overall study-level risk of bias	Main issues
CheckMate 9ER	1L	Low	High	Unclear	Low	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, potential conflict from industry funding
CLEAR	1L	Low	High	Unclear	Unclear	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, very high differential attrition, linked to study endpoints, with methods to account for missing data unclear, some outcomes reported in the trial registry is not reported in the papers (ongoing trial), potential conflict from industry funding
COMPARZ	1L	Low	High	Unclear	Low	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, very high but non-differential attrition with inadequate methods to account for missing data, potential conflict from industry funding
CROSS-J-RCC	1L	Unclear	Unclear	Unclear	Unclear	Unclear	High	**High**	Unclear reporting of randomisation, open-label trial with some subjective outcomes, very high differential attrition with methods to account for missing data unclear, paper reported more outcomes than is listed in the trial registry, unclear conflict as the trial was not industry-funded but some authors received industry funding, risk of carryover effect as no washout period is specified
JAVELIN RENAL 101	1L	Low	High	Unclear	Unclear	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, some imbalances in attrition by reason with inadequate methods to account for missing data, some outcomes reported in the trial registry are not reported in the paper (reported in TA645 but redacted), potential conflict from industry funding

Trial name	Overall line	Selection bias	Performance and detection biases	Attrition bias	Reporting bias	Conflict of interest	Other bias	Overall study-level risk of bias	Main issues
METEOR	2L and 3L	Low	Unclear	Unclear	Low	High	Low	**Unclear**	Open-label trial with some subjective outcomes, very high differential attrition, but linked to study endpoints, with inadequate methods to account for missing data, potential conflict from industry funding
NCT01136733	2L	High	Unclear	High	Low	High	Low	**High**	Dynamic allocation of treatment, small sample with potential baseline imbalances, open-label trial with some subjective outcomes, very high differential attrition, linked to study endpoints as well as other reasons, with inadequate methods to account for missing data, potential conflict from industry funding
RECORD-1	2L and 3L	Low	Low	Unclear	Unclear	High	Low	**Unclear**	Very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, some outcomes reported in the trial registry are not reported in the paper, potential conflict from industry funding
SWITCH	1L	Unclear	Unclear	Unclear	Unclear	High	Unclear	**Unclear**	Unclear reporting of randomisation, open-label trial with some subjective outcomes, some imbalances in attrition by reason with methods to account for missing data unclear, paper reported more outcomes than is listed in the trial registry, potential conflict from industry funding, unclear risk of carryover effect as washout period may be insufficient
SWITCH II	1L	Unclear	Unclear	Unclear	Unclear	High	High	**High**	Unclear reporting of randomisation and allocation concealment, open-label trial with some subjective outcomes, very high but non-differential attrition with methods to account for missing data unclear, paper reported outcomes not listed in the trial registry and did not report other outcomes listed in the trial registry, potential conflict from industry funding, risk of carryover effect as no washout period is specified

Trial name	Overall line	Selection bias	Performance and detection biases	Attrition bias	Reporting bias	Conflict of interest	Other bias	Overall study-level risk of bias	Main issues
TIVO-1	1L and 2L	Low	High	Unclear	Unclear	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, some outcomes reported in the trial registry are not reported in the papers, potential conflict from industry funding
TIVO-3	3L and 4L	Low	Unclear	Unclear	Low	High	Low	**Unclear**	Open-label trial with some subjective outcomes, very high differential attrition, but linked to study endpoints, with inadequate methods to account for missing data, potential conflict from industry funding
De-prioritised									
VEG105192	1L and 2L	Low	Low	Unclear	Unclear	High	Low	**Unclear**	Very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, some outcomes reported in the trial registry are not reported in the paper, potential conflict from industry funding
ASPEN	1L	Unclear	High	Unclear	Low	High	Low	**Unclear**	Unclear reporting of randomisation, open-label trial with some highly subjective outcomes, some imbalances in attrition by reason with methods to account for missing data unclear, potential conflict from industry funding
BIONIKK	1L	High	Unclear	Unclear	Low	High	Low	**High**	Small sample with baseline imbalances, open-label trial with some subjective outcomes, some imbalances in attrition by reason with methods to account for missing data unclear, potential conflict from industry funding
ESPN	1L	Unclear	Unclear	Unclear	Low	High	Low	**Unclear**	Unclear reporting of randomisation and allocation concealment, small sample with potential baseline imbalances, open-label trial with some subjective outcomes, very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, potential conflict from industry funding

Trial name	Overall line	Selection bias	Performance and detection biases	Attrition bias	Reporting bias	Conflict of interest	Other bias	Overall study-level risk of bias	Main issues
Hutson 2017	1L	Unclear	High	Unclear	Low	High	Low	**Unclear**	Unclear reporting of randomisation, open-label trial with some highly subjective outcomes, some imbalances in attrition by reason with methods to account for missing data unclear, potential conflict from industry funding
RECORD-3	1L	Low	High	Unclear	Low	High	Low	**Unclear**	Open-label trial with some highly subjective outcomes, very high but non-differential attrition with inadequate methods to account for missing data, potential conflict from industry funding
SWOG 1500	1L	High	Unclear	Unclear	Low	Unclear	Low	**High**	Dynamic allocation of treatment, small sample with potential baseline imbalances, open-label trial with some subjective outcomes, very high differential attrition, but linked to study endpoints, with methods to account for missing data unclear, unclear conflict as the trial was not industry-funded but some authors received industry funding

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line

Selection bias

Overall, thirteen of the included trials were assessed as having a low risk of selection bias. Only one of these (AXIS) reported fully on an adequate method of random sequence generation, while the remaining 12 described randomisation using interactive voice or web response systems (IxRS). As this has been accepted as evidence of random sequence generation in previous company submissions and NICE TAs, the EAG pragmatically accepted these trials as having adequate methods of sequence generation. However, the EAG views the use of IxRS to be a feature of allocation concealment and considers sequence generation to have been poorly reported in the 12 trials. In the absence of precedent, the EAG would have judged these trials as having an unclear risk of bias relating to sequence generation. Seven of the included trials had an unclear risk of selection bias, driven in large part by unclear descriptions of the methods used to generate the random sequence. Three of these trials additionally did not report adequate allocation concealment in sufficient detail; two trials included small sample sizes and reported some baseline imbalances between randomised groups.

Four trials had a high overall risk of selection bias. This was driven largely by an inadequate randomisation method namely dynamic allocation - a primarily deterministic, non-random approach to balance prognostic factors at baseline[111] in three trials (CABOSUN, NCT01136733 and SWOG 1500). Two of these trials (NCT01136733 and SWOG 1500) had additional potential sources of bias as they included small sample sizes and reported potential baseline imbalances despite the dynamic allocation processes. A fourth trial (BIONIKK), despite describing adequate methods of random sequence generation and allocation concealment in sufficient detail, showed imbalances in most baseline characteristics of participants due to very small sample sizes.

Performance and detection bias

Only two trials were judged as having a low overall risk of performance and detection bias, as they described blinding all groups to treatment assignment. As such, none of the outcomes assessed in these trials were considered to be at risk of these biases. Ten trials were judged to be at unclear risk overall as patients and investigators were not blinded and since both groups were involved in the assessment of outcomes that are, to some extent, subjective. Twelve trials were considered to be at high risk overall as patients and investigators were not blinded and since both groups were involved in the assessment of subjective outcomes.

A lack of blinding in the 22 trials was not considered to have a major impact on OS as this is a hard, objective outcome. In eight trials where the assessors of radiological outcomes based on Response Evaluation Criteria in Solid Tumors (RECIST) criteria were blinded, but other groups were not, outcomes such as PFS, response rate, duration of response, and time to treatment failure were similarly judged as having no major impact due to the blinded and predominantly objective nature of the outcome assessment. Given the largely objective nature of these outcomes, an overall lack of blinding was still considered to be very unlikely to impact on bias. Conversely, a lack of blinding was considered to pose some risk of bias for outcomes such as adverse events, which are patient-reported and, to some extent, investigator-determined; a lack of blinding was judged very likely to result in bias for patient-reported quality of life outcomes.

Attrition bias

Attrition bias was not a major concern with safety outcomes, since most RCTs analysed these in all participants who had received at least one dose of a study treatment. The only exceptions were the BERAT trial, where adverse events were not reported following crossover, and the SWOG 1500 trial, where one patient in each arm received no protocol therapy and were excluded from the safety assessment.

For effectiveness outcomes, 22 and two trials were at unclear and high risk of attrition bias, respectively. Only three trials (CABOSUN, RECORD-3 and SWITCH II) did not report unexpected imbalances in drop-outs between groups. Very high non-differential overall attrition was observed in all three , though reasons for attrition and numbers per reason were similar. However, due to inadequate reporting of appropriate intention-to-treat (ITT) analytical approaches or the reporting of inadequate approaches to account for missing data in the presence of this attrition, all three trials were judged as having unclear risk of attrition bias. Nine additional trials (AXIS, CheckMate 9ER, CLEAR, ESPN, METEOR, NCT01136733, RECORD-1, SWOG1500 and TIVO-1) reported very high, differential dropouts between groups, but these were predominantly linked to study endpoints and were not judged to be unexplained imbalances. However, due to inadequate reporting of appropriate ITT approaches or the reporting of inadequate approaches to account for missing data unrelated to study endpoints, all were judged as having an unclear risk of attrition bias. Eleven trials (ASPEN, BERAT, BIONIKK, CheckMate 025, CheckMate 214, COMPARZ, Hutson 2017, JAVELIN RENAL 101,SWITCH, TIVO-3 and VEG105192) reported high or very high attrition by specific reason not related to study endpoints; with such high attrition it is not possible to rule out that some

drop-outs were related to the true value of the outcomes. This was not mitigated by the description of adequate ITT approaches for ten of the trials and, as a result, these trials were judged also as having an unclear risk of attrition bias. The eleventh trial (BERAT) did not report any ITT approach to dealing with missing data and was judged as having a high risk of attrition bias. Finally, one trial (CROSS-J-RCC) reported very high differential attrition between groups and did not provide reasons for dropouts or discontinuations in the second line. This was not mitigated by the description of adequate ITT approaches and the trial was judged as having an unclear risk of attrition bias.

All trials but one reported using ITT analysis approaches: BERAT did not account for missing data following crossover. For the remaining trials, the appropriateness of the ITT approach either could not be fully assessed due to insufficient detail about how missing data were handled or was found to be lacking. In the case of the latter, protocols or statistical analysis plans seem to indicate last observation carried forward (LOCF) and other single imputation approaches to missing dates. As per Cochrane guidance, these approaches are not considered to be appropriate methods of imputation that address bias, as they are 'unlikely to remove the bias that occurs when missingness in the outcome depends on its true value, unless there is no change in the outcome after the last time it was measured'.[112]

Reporting bias

Fifteen trials did not have evidence of reporting bias, as all outcomes listed in trial registries or published protocols were reported on. Two exceptions were the BIONIKK trial, where certain laboratory outcomes listed were not yet reported on, but intentions to do so separately were reported; and for the CABOSUN and TIVO-3 trials, where adverse events were not listed but reported – this was considered a reasonable inclusion and not a source of potential bias.

Nine trials had unclear risk of reporting bias as the publications reported either more or less outcomes than those listed in the trial registry or protocol. In terms of specific outcomes of interest, it was noted that the CLEAR trial is ongoing and may still be measuring the relevant outcomes, though a recent American Society of Clinical Oncology conference abstract[109] provided no additional data on these outcomes. JAVELIN RENAL 101 did not report EQ-5D and FKSI results in the publications, but did provide these data in NICE TA645; however, these are redacted and not available to the EAG. Furthermore, TIVO-1 listed duration of response in its trial registry but did not report this outcome in the publications.

Conflicts of interest

Only two studies were considered to be at unclear risk for conflict of interest: both were not sponsored by pharmaceutical companies, though authors did list the receipt of various funds from pharmaceutical companies.

Authors of all remaining trials reported receiving fees, grants and other monies from pharmaceutical companies, including the company that sponsored the trial; in several cases some authors also declared being employees or holding stock in the sponsoring company. These trials were considered to have a high risk of bias related to conflict of interest.

Other biases

No specific other biases were identified for the parallel trials. Three crossover trials (BERAT, CROSS-J-RCC and SWITCH-II) were identified as having high risk of bias due to carryover effects as no washout period was specified; therefore, post-crossover results should be interpreted with caution. Another crossover trial (SWITCH) did specify a washout period, but it was not clear whether this was of sufficient duration to eliminate all carryover effects. This trial was judged to have unclear bias for this domain. The two remaining crossover trials (ESPN and RECORD-3) were considered to be at low risk of bias due to carryover effects as the washout periods specified were longer than the clearance of the included treatments.

Overall bias

The overall bias of the included trials, assessed by considering the worst-rated of the key domains (parallel RCTs: selection and attrition bias; crossover RCTs: selection, attrition and other bias) as the overall judgment, indicated that none of the trials were at low overall risk of bias. According to this approach, seventeen trials were judged to be at unclear overall risk of bias; all were at unclear risk for the attrition bias domain while four were additionally at unclear risk of selection bias. Eleven trials were judged to be at high overall risk of bias, primarily due to a high risk of attrition bias. Overall attrition in the various arms of trials judged to be at high risk of attrition bias (BERAT and NCT01136733,) ranged from 20% to 94%. Five of the trials at high overall risk of bias did not have a high risk of bias for the attrition domain; instead, overall bias for these trials was driven by a high risk of selection bias, due to dynamic allocation (CABOSUN and SWOG 1500) or considerable baseline imbalances (BIONIKK), or a high risk of bias due to carryover effects in two crossover trials (CROSS-J-RCC and SWITCH II). Two trials had more than one key domain at high risk of bias, with BERAT at high risk of both attrition bias and bias

relating to carryover effect following crossover and NCT01136733 at high risk of both selection and attrition biases.

CheckMate 9ER, the key trial of interest, was judged to have an unclear overall risk of bias because of an unclear risk of attrition bias. Very high, differential overall attrition (44% in the cabozantinib + nivolumab (CABO/NIV) arm and 71% in the sunitinib (SUN) arm) was reported; however, this related to study endpoints with considerable dropouts due to discontinuation (43% CABO/NIV and 69% SUN) and disease progression (27% CABO/NIV and 46% SUN). The reporting of single imputation approaches was not considered an ideal method to deal with missing data unrelated to study outcomes.. Random sequence generation was poorly reported, but pragmatically accepted as presenting low risk of bias due to the use of IxRS for randomisation.

Consequently, results from the NMA are based on underlying evidence with various methodological shortcomings. Most notable of these is very high attrition with inadequate or unclear approaches to handling missing data and demonstrating that missingness is the outcome is not related to its true value. It is highly unlikely that such high attrition would not effectively subvert randomisation as missingness is likely to depend on the true value of the outcome.

3.3.3. Clinical effectiveness results from trials identified in the review

3.3.3.1. Overall survival

1st line

Overall risk

Nine prioritised trials evaluated OS in an overall risk population in the 1st line setting. All trials included a comparison with sunitinib (7 trials) and/or sorafenib (four trials). Two trials compared sunitinib and sorafenib and found no clear difference in OS between the two treatments. Pazopanib was evaluated in two trials, otherwise all interventions (avelumab + axitinib; tivozanib, cabozantinib + nivolumab, pembrolizumab + lenvatinib, and nivolumab + ipilimumab) were evaluated in only one trial. There was no clear difference between pazopanib and either sunitinib or sorafenib. Median OS was highly variable for sunitinib, ranging between 27.4 – 54.3 months. Median OS was between 29.3 – 30 months for sorafenib and was 28.3 for pazopanib.

Cabozantinib + nivolumab and nivolumab + ipilimumab were associated with the largest benefits for OS compared with sunitinib (CheckMate 9ER and CheckMate 214). These were followed by pembrolizumab + lenvatinib in the CLEAR trial, though 95% confidence intervals around the effect reached the line of null effect. It was noted, however, that median PFS in the sunitinib arm of CLEAR was significantly greater than in either CheckMate 9ER or CheckMate 214 (54.3 months compared to 35.5 and 38.4 months). The EAG did not identify a clear reason for the difference between trials. Median OS had not been reached in the latest data cut for avelumab + axitinib, though initial findings suggest that this performed well in comparison to sunitinib. There was no benefit for tivozanib over sorafenib.

Favourable risk

Seven trials reported OS at 1st line for the favourable risk group. All trials involved a comparison with sunitinib (7 trials) and/or sorafenib (2 trials). Median OS was not reached or not reported for most trials, though where available median OS ranged from 43.6 to 68.4 months for sunitinib. The other treatments (nivolumab + ipilimumab, pembrolizumab + lenvatinib, avelumab + axitinib,

cabozantinib + nivolumab, and pazopanib) were each evaluated in only one trial. All relative effects were associated with extremely wide 95% confidence intervals, largely due to the small sample size and the lack of available data at the time of calculation. As a consequence of this and unexplained variation between trials, no treatment was clearly associated with a clinical benefit for OS over its comparator.

Intermediate/poor risk

Eight trials reported OS at 1st line in an intermediate/poor risk population. All trials involved a comparison with sunitinib (8 trials). Sorafenib was only compared with sunitinib (2 trials). All other treatments (nivolumab + ipilimumab, pembrolizumab + lenvatinib, avelumab + axitinib, pazopanib, cabozantinib, and cabozantinib + nivolumab) were each evaluated by only one trial. Median OS ranged between 21.2 – 37.8 months for sunitinib (NR for sorafenib). A clinical benefit was seen for both nivolumab + ipilimumab and cabozantinib + nivolumab in comparison with sunitinib. A benefit was also seen for pembrolizumab + lenvatinib and avelumab + axitinib in comparison with sunitinib, though in both cases the 95% confidence intervals approached the line of null effect. A benefit was seen for cabozantinib in CABOSUN, though this was the trial with the smallest number of participants (n=158) and 95% confidence intervals spanned widely both sides of the line of null effect and median OS was considerably shorter than was reported for other interventions. Median OS for nivolumab + ipilimumab, cabozantinib + nivolumab, pembrolizumab + lenvatinib, and avelumab + axitinib all exceeded 40 months.

2nd line-plus

Seven trials reported OS in the 2nd line setting, all in an overall risk population. Everolimus was evaluated in five trials, sorafenib and axitinib were each evaluated in two trials, and all other treatments (nivolumab, cabozantinib, everolimus + lenvatinib, tivozanib and placebo) were each evaluated in one trial. Median OS following everolimus was fairly consistent across trials, ranging from 15.3 to 16.5 months. Cabozantinib, nivolumab, and everolimus + lenvatinib all outperformed everolimus alone. There was no clear difference between everolimus, sorafenib, axitinib, and tivozanib.

Table 13: OS in prioritised included trials

Trial name	First author	Intervention name	Control name	Risk group	Follow-up time category	N (int)	N (control)	Median OS (95%CI)	HR (95%CI)
1L									
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	Overall	5yr+	550	546	Int: 55.7 (NR); Control: 38.4 (NR)	0.72 (0.62, 0.85)
CLEAR	Motzer (2023)	Pem+lenv	Suni	Overall	4-5yr	355	357	Int: 53.7 (48.7, NE); Control: 54.3 (40.9, NE)	0.79 (0.63, 0.99)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Overall	4-5yr	60	64	Int: 38.4 (NR); Control: 30.9 (NR)	0.934 (0.588, 1.485)
SWITCH	Eichelberg (2015)	Sora	Suni	Overall	1-2yr	182	183	Int: 30 (NR); Control: 27.4 (NR)	0.99 (0.73, 1.33)
SWITCH II	Retz (2019)	Sora	Pazo	Overall	3-4yr	189	188	Int: NR; Control: NR	1.22 (0.91, 1.65)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	Overall	2-3yr	442	444	Int: NE (42.2, NE); Control: 37.8 (31.4, NE)	0.79 (0.64, 0.97)
COMPARZ	Motzer (2014)	Pazo	Suni	Overall	2-3yr	557	553	Int: 28.3 (26.0, 35.5); Control: 29.1 (25.4, 33.1)	0.92 (0.79, 1.06)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	Overall	3-4yr	323	328	Int: 49.48 (40.31, NE); Control: 35.52 (29.24, 42.25)	0.7 (0.56, 0.87)
TIVO-1	Motzer (2013)	Tivo	Sora	Overall	2-3yr	260	257	Int: 29.3 (NR); Control: 28.8 (NR)	1.245 (0.95, 1.62)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	Fav	5yr+	125	124	Int: 74.1 (NR); Control: 68.4 (NR)	0.94 (0.65, 1.37)
CLEAR	Motzer (2023)	Pem+lenv	Suni	Fav	4-5yr	110	124	Int: Not reached (NR); Control: 59.9 (58.8, NE)	0.94 (0.58, 1.52)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Fav	4-5yr	12	14	Int: NR; Control: NR	0.35 (0.1, 1.2)
SWITCH	Eichelberg (2015)	Sora	Suni	Fav	1-2yr	71	82	Int: NR; Control: NR	1.24 (0.61, 2.56)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	Fav	2-3yr	94	96	Int: NE (NE, NE); Control: NE (39.8, NE)	0.66 (0.36, 1.22)

COMPARZ	Motzer (2014)	Pazo	Suni	Fav	2-3yr	151	152	Int: 42.5 (37.9, NE); Control: 43.6 (37.1, 47.4)	0.88 (0.63, 1.21)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	Fav	3-4yr	74	72	Int: NE (40.67, NE); Control: 47.61 (43.63, NE)	1.07 (0.63, 1.79)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	Int/poor	5yr+	425	422	Int: 47 (NR); Control: 26.6 (NR)	0.68 (0.58, 0.81)
CLEAR	Motzer (2023)	Pem+lenv	Suni	Int/poor	4-5yr	243	229	Int: 47.9 (40.5, NE); Control: 34.3 (26.3, 54.3)	0.74 (0.57, 0.96)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Int/poor	4-5yr	45	49	Int: NR; Control: NR	1.2 (0.7, 1.95)
SWITCH	Eichelberg (2015)	Sora	Suni	Int/poor	1-2yr	108	94	Int: NR; Control: NR	0.83 (0.53, 1.31)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	Int/poor	2-3yr	343	347	Int: 42.2 (33.1, NE); Control: 37.8 (29.6, NE)	0.79 (0.64, 0.98)
COMPARZ	Motzer (2014)	Pazo	Suni	Int/poor	2-3yr	389	380	Int: NR; Control: NR	0.891 (0.75, 1.06)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	Int/poor	3-4yr	249	256	Int: 49.5 (34.9, NE); Control: 29.2 (23.7, 36.0)	0.65 (0.51, 0.83)
CABOSUN	Choueiri (2018)	Cabo	Suni	Int/poor	2-3yr	79	78	Int: 26.6 (14.6, NE); Control: 21.2 (16.3, 27.4)	0.8 (0.53, 1.21)
2L+									
AXIS	Motzer (2013)	Axi	Sora	Overall	3-4yr	361	362	Int: 20.1 (16.7, 23.4); Control: 19.2 (17.5, 22.3)	0.969 (0.8, 1.174)
BERAT	Grunwald (2022)	Evero	Axi	Overall	1-2yr	5	5	Int: 15.29 (6.0, NE); Control: 18.64 (5.9, 32.5)	1.12 (0.27, 4.61)
CheckMate 025	Escudier (2022)	Nivo	Evero	Overall	5yr+	410	411	Int: NR; Control: NR	0.74 (0.63, 0.86)
METEOR	Choueiri (2016)	Cabo	Evero	Overall	1-2yr	330	328	Int: 21.4 (18.7, NE); Control: 16.5 (14.7, 18.8)	0.66 (0.53, 0.83)
NCT01136733	Motzer (2015)	Lenv+evero	Evero	Overall	2-3yr	51	50	Int: 25.5 (16.4, NE); Control: 15.4 (11.8, 19.6)	0.51 (0.3, 0.88)
RECORD-1	Motzer (2010)	Evero	PBO	Overall	1-2yr	277	139	Int: 14.8 (NR); Control: 14.4 (NR)	0.87 (0.65, 1.15)

TIVO-3	Rini (2022)	Tivo	Sora	Overall	1-2yr	175	175	Int: NR; Control: NR	0.89 (0.7, 1.14)

Abbreviations: axi, axitinib; cabo, cabozantinib; CI; confidence interval; evero, everolimus; HR, hazard ratio; int, intermediate; ipi, ipilimumab; lenv, lenvatinib; NE, not estimable; nivo, nivolumab; NR, not reported; OS, overall survival; pazo, pazopanib; PBO, placebo; pem, pembrolizumab; sora, sorafenib; suni, sunitinib; tivo, tivozanib

3.3.3.2. Progression-free survival

1st line

Overall risk

Nine trials reported PFS for the overall risk population in the 1st line setting. All trials involved a comparison either with sunitinib or sorafenib. Sunitinib outperformed sorafenib: median PFS ranged across trials as 5.6 – 9.1 months for sorafenib and 8.4 – 10.2 months for sunitinib. Pazopanib was evaluated in two trials, while all other treatments were evaluated in one trial only. Pazopanib outperformed sorafenib but was no different to sunitinib. In order of best performing treatments first, the treatments that performed better than sunitinib were pembrolizumab + lenvatinib, cabozantinib + nivolumab, avelumab + axitinib, and nivolumab + ipilimumab.

Favourable risk

In the favourable risk group, eight trials reported PFS in the 1st line setting. All trials involved a comparison either with sunitinib or sorafenib. Sunitinib outperformed sorafenib: no trials reported median PFS for sorafenib, while two trials reported median PFS for sunitinib as 13.8 – 13.9 months. All other treatments were evaluated in one trial only. Sunitinib outperformed nivolumab + ipilimumab. In order of best performing treatment first, pembrolizumab + lenvatinib, tivozanib, avelumab + axitinib, and cabozantinib + nivolumab outperformed sunitinib. However, in the case of avelumab + axitinib and cabozantinib + nivolumab, 95% confidence intervals crossed the line of null effect, suggesting some meaningful uncertainty in the findings. There was no difference between pazopanib and sunitinib.

Intermediate/poor risk

In the intermediate/poor risk group, nine trials evaluated PFS in the 1st line setting. All trials involved a comparison either with sunitinib or sorafenib. There was no clear difference in PFS between sunitinib and sorafenib. All other treatments were evaluated in one trial only. There was no difference between pazopanib and sunitinib. In order of best performing treatments first, the treatments that performed better than sunitinib or sorafenib were pembrolizumab + lenvatinib, cabozantinib, cabozantinib + nivolumab,

avelumab + axitinib, nivolumab + ipilimumab, and tivozanib. For tivozanib, 95% confidence intervals crossed the line of null effect and there was therefore meaningful uncertainty in this result.

It was noted that while cabozantinib + nivolumab performed similarly to cabozantinib alone in comparison with sunitinib in the intermediate/poor risk group, median PFS was longer for cabozantinib + nivolumab than for cabozantinib alone. There were differences between trials that could reduce the comparability of effects between trials; CABOSUN was noted to be a smaller trial set in the USA only, and with a slightly higher rate of participants with bone metastases. However, given the magnitude of difference in the median PFS between cabozantinib and cabozantinib + nivolumab, the EAG considered it plausible that the addition of nivolumab was associated with an increased benefit over sunitinib than cabozantinib alone. Further evidence may be needed to resolve the extent of this benefit.

2nd line-plus

In the 2nd line setting, eight trials evaluated PFS, all in an overall risk population. The treatments evaluated were everolimus (five trials), cabozantinib (1 trial), everolimus + lenvatinib (1 trial), sorafenib (3 trials) tivozanib (2 trials), nivolumab (1 trial), axitinib (1 trial) and placebo (1 trial). All trials included a comparison with either placebo, everolimus or sorafenib. Median PFS was 1.9 months for placebo, between 3.7 to 5.5 months for everolimus, and was 3.9 to 5.7 for sorafenib. The longest PFS was reported for everolimus + lenvatinib at 14.6 months, though there was considerable uncertainty in this (95% Cis 5.9, 20.1). Cabozantinib, everolimus + lenvatinib and nivolumab each out-performed everolimus alone, though the effect of nivolumab was uncertain due to imprecision. Axitinib was shown to outperform sorafenib, as did tivozanib though with some uncertainty.

Table 14: PFS in prioritised included trials

Trial name	Author (year)	Int. name	Cont. name	Risk group	BICR / IA	PFS assessment method	Follow-up time cat.	N (int)	N (control)	Median PFS (95%CI)	HR (95%CI)
1L											
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	Overall	IA	RECIST, FDA rule	5yr+	550	546	Int: NR; Control: NR	0.86 (0.73, 1.01)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	Overall	BICR	RECIST, FDA rule	3-4yr	323	328	Int: 16.6 (12.8, 19.5); Control: 8.4 (7.0, 9.7)	0.59 (0.49, 0.71)
CLEAR	Motzer (2023)	Pem+lenv	Suni	Overall	ICR (no blinding)	RECIST, FDA rule	4-5yr	355	357	Int: 23.9 (20.8, 27.7); Control: 9.2 (6.0, 11.0)	0.47 (0.38, 0.57)
COMPARZ	Motzer (2013)	Pazo	Suni	Overall	BICR	RECIST, FDA rule	1-2yr	557	553	Int: 8.4 (8.3, 10.9); Control: 9.5 (8.3, 11.1)	1.05 (0.9, 1.22)
COMPARZ	Motzer (2013)	Pazo	Suni	Overall	IA	RECIST, FDA rule	1-2yr	557	553	Int: 10.5 (8.3, 11.1); Control: 10.2 (8.3, 11.1)	1 (0.86, 1.15)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Overall	IA	RECIST, censoring rules unclear	4-5yr	60	64	Int: 8.7 (5.5, 21.1); Control: 7 (6.1, 12.2)	0.67 (0.42, 1.08)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	Overall	BICR	RECIST, FDA rule	2-3yr	442	444	Int: 13.9 (11.1, 16.6); Control: 8.5 (8.2, 9.7)	0.67 (0.57, 0.79)
SWITCH	Eichelberg (2015)	Sora	Suni	Overall	IA	RECIST, FDA rule	<1yr	182	183	Int: 5.9 (NR); Control: 8.5 (NR)	1.19 (0.93, 1.53)
SWITCH II	Retz (2019)	Sora	Pazo	Overall	IA	RECIST, censoring rules unclear	3-4yr	189	188	Int: 5.6 (4.7, 6.3); Control: 9.3 (7.4, 10.6)	1.51 (1.19, 1.92)

Trial name	Author (year)	Int. name	Cont. name	Risk group	BICR / IA	PFS assessment method	Follow-up time cat.	N (int)	N (control)	Median PFS (95%CI)	HR (95%CI)
TIVO-1	Motzer (2013)	Tivo	Sora	Overall	BICR	RECIST, FDA rule	NR	181	181	Int: 12.7 (NR); Control: 9.1 (NR)	0.76 (0.58, 0.99)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	Fav	IA	RECIST, FDA rule	RECIST, FDA rule	125	124	Int: NR; Control: NR	1.6 (1.13, 2.26)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	Fav	BICR	RECIST, FDA rule	3-4yr	74	72	Int: 21.42 (13.08, 24.71); Control: 13.86 (9.56, 16.66)	0.72 (0.49, 1.05)
CLEAR	Motzer (2023)	Pem+lenv	Suni	Fav	ICR (no blinding)	RECIST, FDA rule	4-5yr	110	124	Int: NR; Control: NR	0.5 (0.35, 0.71)
COMPARZ	Motzer (2013)	Pazo	Suni	Fav	BICR	RECIST, FDA rule	1-2yr	151	152	Int: NR; Control: NR	1.01 (0.74, 1.37)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Fav	IA	RECIST, censoring rules unclear	4-5yr	12	14	Int: NR; Control: NR	0.245 (0.082, 0.734)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	Fav	BICR	RECIST, FDA rule	2-3yr	94	96	Int: 20.7 (16.6, 26.3); Control: 13.8 (11.1, 23.5)	0.71 (0.49, 1.016)
SWITCH	Eichelberg (2015)	Sora	Suni	Fav	IA	RECIST, FDA rule	<1yr	71	82	Int: NR; Control: NR	1.3 (0.81, 2.09)
TIVO-1	Motzer (2013)	Tivo	Sora	Fav	BICR	RECIST, FDA rule	NR	70	87	Int: NR; Control: NR	0.59 (0.378, 0.921)
CABOSUN	Choueiri (2018)	Cabo	Suni	Int/poor	BICR	RECIST, FDA rule	2-3yr	79	78	Int: 8.6 (6.8, 14.0); Control: 5.3 (3.0, 8.2)	0.48 (0.31, 0.74)

Trial name	Author (year)	Int. name	Cont. name	Risk group	BICR / IA	PFS assessment method	Follow-up time cat.	N (int)	N (control)	Median PFS (95%CI)	HR (95%CI)
CABOSUN	Choueiri (2018)	Cabo	Suni	Int/poor	IA	RECIST, FDA rule	2-3yr	79	78	Int: 8.3 (6.5, 12.4); Control: 5.4 (3.4, 8.2)	0.56 (0.37, 0.83)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	Int/poor	BICR	RECIST, FDA rule	5yr+	425	422	Int: NR; Control: NR	0.73 (0.61, 0.87)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	Int/poor	BICR	RECIST, FDA rule	RECIST, FDA rule	249	256	Int: 15.61 (11.17, 19.15); Control: 7.05 (5.68, 8.90)	0.56 (0.46, 0.69)
CLEAR	Motzer (2023)	Pem+lenv	Suni	Int/poor	ICR (no blinding)	RECIST, FDA rule	4-5yr	243	229	Int: NR; Control: NR	0.43 (0.34, 0.55)
COMPARZ	Motzer (2013)	Pazo	Suni	Int/poor	BICR	RECIST, FDA rule	1-2yr	322	328	Int: NR; Control: NR	0.98 (0.80, 1.19)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Int/poor	IA	RECIST, censoring rules unclear	4-5yr	45	49	Int: NR; Control: NR	1 (0.62, 1.63)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	Int/poor	BICR	RECIST, FDA rule	2-3yr	343	347	Int: 12.9 (11.1, 16.6); Control: 8.4 (7.9, 10.1)	0.66 (0.55, 0.787)
SWITCH	Eichelberg (2015)	Sora	Suni	Int/poor	IA	RECIST, FDA rule	<1yr	108	94	Int: NR; Control: NR	1.14 (0.77, 1.67)
TIVO-1	Motzer (2013)	Tivo	Sora	Int/poor	BICR	RECIST, FDA rule	NR	190	170	Int: NR; Control: NR	0.821 (0.635, 1.062)
2L+											
AXIS	Motzer (2013)	Axi	Sora	Overall	BICR	RECIST, censoring rules unclear	3-4yr	361	362	Int: 8.3 (6.7, 9.2); Control: 5.7 (4.7, 6.5)	0.66 (0.55, 0.78)

Trial name	Author (year)	Int. name	Cont. name	Risk group	BICR / IA	PFS assessment method	Follow-up time cat.	N (int)	N (control)	Median PFS (95%CI)	HR (95%CI)
BERAT	Grunwald (2022)	Evero	Axi	Overall	IA	RECIST, censoring rules unclear	1-2yr	5	5	Int: 3.7 (2.6, 8.4); Control: 2.2 (1.9, NC)	1 (0.26, 3.85)
CheckMate 025	Escudier (2022)	Nivo	Evero	Overall	IA	RECIST, FDA rule	5yr+	410	411	Int: NR; Control: NR	0.84 (0.72, 0.99)
METEOR	Choueiri (2016)	Cabo	Evero	Overall	BICR	RECIST, FDA rule	<1yr	330	328	Int: 7.4 (6.6, 9.1); Control: 3.9 (3.7, 5.1)	0.51 (0.41, 0.62)
METEOR	Choueiri (2016)	Cabo	Evero	Overall	IA	RECIST, FDA rule	<1yr	330	328	Int: 7.4 (6.6, 9.1); Control: 5.1 (3.9, 5.5)	0.54 (0.44, 0.65)
NCT01136733	Motzer (2015)	Lenv +evero	Evero	Overall	IA	RECIST, censoring rules unclear	1-2yr	51	50	Int: 14.6 (5.9, 20.1); Control: 5.5 (3.5, 7.1)	0.4 (0.24, 0.68)
RECORD-1	Motzer (2010)	Evero	PBO	Overall	BICR	RECIST, FDA rule	1-2yr	277	139	Int: 4.9 (4.0, 5.5); Control: 1.9 (1.8, 1.9)	0.33 (0.25, 0.43)
RECORD-1	Motzer (2010)	Evero	PBO	Overall	IA	RECIST, FDA rule	1-2yr	277	139	Int: 5.5 (4.6, 5.8); Control: 1.9 (1.8, 2.2)	0.32 (0.25, 0.41)
TIVO-3	Atkins (2022)	Tivo	Sora	Overall	IA	RECIST, FDA rule	1-2yr	175	175	Int: NR; Control: NR	0.624 (0.49, 0.79)
TIVO-3	Rini (2020)	Tivo	Sora	Overall	BICR	RECIST, FDA rule	1-2yr	175	175	Int: 5.6 (5.29, 7.33); Control: 3.9 (3.71, 5.55)	0.73 (0.56, 0.94)

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; CI; confidence interval; evero, everolimus; FDA, Food and Drug Administration; HR, hazard ratio; int, intermediate; ipi, ipilimumab; lenv, lenvatinib; NE, not estimable; nivo, nivolumab; NR, not reported; OS, overall survival; pazo, pazopanib; PBO, placebo; pem, pembrolizumab; PFS, progression-free survival; RECIST, Response Evaluation Criteria in Solid Tumours sora, sorafenib; suni, sunitinib; tivo, tivozanib

3.3.3.3. Response rates

1st line

Overall risk

Nine trials reported response rates in an overall risk population at 1st line. All trials involved either a comparison with sunitinib (9 trials) and/or sorafenib (3 trials). All other treatments (pazopanib, avelumab + axitinib, cabozantinib + nivolumab, pembrolizumab + lenvatinib, and nivolumab + ipilimumab) were evaluated in one trial only.

Response rates for sunitinib ranged between 23.3% and 36.8%. There was a trend for response rates to increase slightly with longer follow-up, with some exceptions. Response rates for sorafenib across trials ranged from 15.6% to 30.2%, with no pattern related to follow-up duration. Two trials compared sunitinib and sorafenib and did not find any clear difference in response rate.

Large effects were reported for (in order of best performing treatments first) pembrolizumab + lenvatinib, cabozantinib + nivolumab, and avelumab + axitinib, all in comparison with sunitinib. A moderate benefit was also reported for nivolumab + ipilimumab in comparison with sunitinib.

Favourable risk

Four trials reported response rate in a favourable risk population in the 1st line. All trials involved a comparison with sunitinib. Response rates for sunitinib ranged between 45.8% to 52%, with no trend over time. In order of the best performing treatments first, large effects were seen for avelumab + axitinib and cabozantinib + nivolumab, and a moderate effect for pembrolizumab + lenvatinib. A lower rate of response was shown following nivolumab + ipilimumab in comparison with sunitinib.

Intermediate/poor risk

Five trials reported response rates in an intermediate/poor risk population in the 1st line. All trials involved a comparison with either sunitinib (5 trials) or sorafenib (1 trial). All other treatments (cabozantinib, cabozantinib and nivolumab, avelumab + axitinib, nivolumab + ipilimumab, and tivozanib) were evaluated in only one trial.

Response rates for sorafenib were variable across trials, and ranged between 9.0% and 28.8%, with no trend over time. Response rates for sorafenib were reported using both BICR and IA in the TIVO-1 trial, with a difference in response depending on the method used: 23.3% using BICR and 30.7% using IA. A difference in response rate between IA and BICR assessment was also shown for the CABOSUN trial (cabozantinib vs. sunitinib). In general, in other population groups, there was a trend across trials for response rates to be slightly higher when assessed using IA than BICR, though the difference was not universal and not always as large.

A very large effect was reported for pembrolizumab + lenvatinib in comparison with sunitinib, and while the 95% confidence intervals around the effect were large, the lower bounds were still greater than any other reported effect. Large effects were also reported for cabozantinib + nivolumab, avelumab + axitinib, cabozantinib and nivolumab + ipilimumab.

2nd line-plus

Seven trials reported response rates in the 2nd line-plus, all in an overall risk population. Treatments evaluated were everolimus (five trials), sorafenib (two trials), axitinib (2 trials), cabozantinib (1 trial), everolimus + lenvatinib (1 trial), tivozanib (1 trial), nivolumab (1 trial) and placebo (1 trial). Response rates for everolimus and axitinib were fairly consistent across trials: response rates for everolimus were low and ranged between 0% and 6%.

The largest effect was reported for everolimus + lenvatinib in comparison with everolimus alone (a response rate of 43.1% vs 6.0%). Large effects were also reported for cabozantinib and nivolumab. Moderate effects were seen for tivozanib and axitinib.

Table 15: Response rates in prioritised included trials

Trial name	Author (date)	Intervention	Control	Follow-up time category	Risk group	Assessor (IA or BICR)	N (int)	N (control)	Prop (int)	Prop (control)	OR (95%CI)
1L											
SWITCH	Eichelberg (2015)	Sora	Suni	<1yr	Overall	IA	182	183	30.22%	27.87%	1.12 (0.71, 1.76)
COMPARZ	Motzer (2013)	Pazo	Suni	1-2yr	Overall	BICR	557	553	30.70%	24.77%	1.35 (1.03, 1.75)
COMPARZ	Motzer (2013)	Pazo	Suni	1-2yr	Overall	IA	557	553	33.39%	28.93%	1.23 (0.95, 1.59)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	2-3yr	Overall	IA	442	444	59.30%	31.80%	3.13 (2.37, 4.12)
SWITCH II	Retz (2019)	Sora	Pazo	3-4yr	Overall	IA	189	188	28.57%	46.28%	0.46 (0.30, 0.71)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	3-4yr	Overall	BICR	323	328	56.04%	28.05%	3.27 (2.36, 4.53)
CLEAR	Motzer (2023)	Pem+lenv	Suni	4-5yr	Overall	BICR	355	357	71.30%	36.70%	4.28 (3.12, 5.86)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	4-5yr	Overall	Unclear	60	64	23.33%	15.63%	1.64 (0.67, 4.05)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	5yr+	Overall	BICR	550	546	39%	32.00%	1.36 (1.06, 1.74)
CLEAR	Grunwald (2021)	Pem+lenv	Suni	2-3yr	Fav	BICR	74	72	68.20%	50.80%	1.97 (1.01, 3.86)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	2-3yr	Fav	IA	94	96	75.50%	45.80%	3.65 (1.97, 6.77)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	3-4yr	Fav	BICR	74	72	67.57%	45.83%	[redacted]
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	5yr+	Fav	BICR	125	124	30.00%	52.00%	0.41 (0.24, 0.69)
CLEAR	Grunwald (2021)	Pem+lenv	Suni	2-3yr	Int/poor	BICR	188	188	72.40%	28.80%	6.51 (4.15, 10.20)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	2-3yr	Int/poor	IA	343	347	55.10%	28.00%	3.16 (2.30, 4.34)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	3-4yr	Int/poor	BICR	249	256	52.61%	23.05%	[redacted]

Trial name	Author (date)	Intervention	Control	Follow-up time category	Risk group	Assessor (IA or BICR)	N (int)	N (control)	Prop (int)	Prop (control)	OR (95%CI)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	5yr+	Int/poor	BICR	425	422	42.00%	27.00%	1.97 (1.47, 2.62)
CABOSUN	Choueiri (2018)	Cabo	Suni	2-3yr	Overall/int/poor	BICR	79	78	20.25%	8.97%	2.58 (1.00, 6.67)
CABOSUN	Choueiri (2018)	Cabo	Suni	2-3yr	Overall/int/poor	IA	79	78	32.91%	11.54%	3.76 (1.63, 8.70)
TIVO-1*	Motzer (2013)	Tivo	Sora	NR	Overall	BICR	260	257	33.10%	23.30%	1.62 (1.10, 2.39)
TIVO-1*	Motzer (2013)	Tivo	Sora	NR	Overall	IA	260	257	35.40%	30.70%	1.23 (0.85, 1.78)
2L+											
METEOR	Choueiri (2016)	Cabo	Evero	<1yr	Overall	BICR	330	328	17.27%	3.35%	6.02 (3.09, 11.71)
METEOR	Choueiri (2016)	Cabo	Evero	<1yr	Overall	IA	330	328	23.64%	4.27%	6.94 (3.84, 12.56)
AXIS	Rini (2011)	Axi	Sora	1-2yr	Overall	BICR	361	362	19.39%	9.39%	2.32 (1.50, 3.60)
NCT01136733	Motzer (2015)	Lenv+evero	Evero	1-2yr	Overall	IA	51	50	43.14%	6.00%	11.89 (3.26, 43.26)
RECORD-1	Motzer (2010)	Evero	Placebo	1-2yr	Overall	BICR	277	139	1.81%	0.00%	5.63 (0.31, 102.6)
TIVO-3	Verzoni (2021)	Tivo	Sora	1-2yr	Overall	IA	175	175	23.43%	11.43%	2.37 (1.32, 4.25)
AXIS	Motzer (2013)	Axi	Sora	3-4yr	Overall	IA	361	362	22.71%	12.43%	2.07 (1.39, 3.08)
CheckMate 025	Motzer (2020)	Nivo	Evero	5yr+	Overall	IA	410	411	22.93%	4.14%	6.89 (4.03, 11.80)
BERAT	Grunwald (2022)	Evero	Axi	NR ('short')	Overall	IA	5	5	0.00%	20.00%	0.27 (0.01, 8.46)

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; BICR, blinded independent central review; CI; confidence interval; evero, everolimus; IA, investigator-assessed; ipi, ipilimumab; lenv, lenvatinib; NE, not estimable; nivo, nivolumab; NR, not reported; OR, odds ratio; pazo, pazopanib; PBO, placebo; pem, pembrolizumab; PFS, progression-free survival; RECIST, Response Evaluation Criteria in Solid Tumours sora, sorafenib; suni, sunitinib; tivo, tivozanib

* 1L and 2L

3.3.3.4. Duration of response

In total, 9 trials reported the duration of response with treatment, five[106,109,110,113] in the 1st line setting and four[85,97,107,114] in the 2nd line-plus setting.

1st line

In the 1st line population, the comparator in all trials was sunitinib. The median duration of response for sunitinib ranged between 14.5 – 32.0 months for patients at overall risk (5 studies[106,109,110,113]) and was 20.8 months and 9.8 months in those with favourable and poor risk, respectively (1 trial[113]). Duration of response with sunitinib was particularly long for the CheckMate-214 trial compared to the other trials, which did not appear to be explained by the follow-up duration, treatment dose or participant characteristics.

Duration of response was available for avelumab + axitinib in the overall, favourable and intermediate/poor risk groups (1 trial[113]), and cabozantinib + nivolumab (1 trial[110]), pembrolizumab + lenvatinib (1 trial[109]) and NIVO/IPI (1 trial[106]) in the overall risk group. In the overall risk population, and in descending order, median duration of response was not reached for nivolumab + ipilimumab (with a follow-up of over 5 years in CheckMate 214[106]), 26.7 months for pembrolizumab + lenvatinib, 22.08 months for cabozantinib + nivolumab, and 19.4 months for avelumab + axitinib. In the JAVELIN trial,[113] unlike for sunitinib where there was a difference in duration of response between favourable and intermediate/poor risk groups, median duration of response was similar: 22.6 months and 19.3 months for favourable and intermediate/poor risk groups, respectively.

2nd line-plus

In the 2nd line population, all trials reported the duration of response in the overall risk group. Two trials used everolimus[97,107] as the comparator and two trials[85,114] used sorafenib. Median duration of response ranged from 8.5 to 14 months for everolimus and 9 to 10.6 months for sorafenib. A comparison of the two trials using everolimus as a comparator did not satisfactorily resolve the difference in duration of response: while NCT01136733 included a higher proportion of participants at poor risk, it also primarily

included people treated at 2nd line, while more than a quarter of participants in CheckMate 025 (28%) were receiving 3rd line treatment.

Duration of response was available for axitinib (1 trial), lenvatinib + everolimus (1 trial), tivozanib (1 trial) and nivolumab (1 trial). In descending order, median duration of response was 20.3 months for tivozanib, 18.2 months for nivolumab, 13 months for lenvatinib + everolimus, and 11 months for axitinib.

Table 16: Duration of response in prioritised included trials

Trial name	First author	Int. name	Control name	Follow-up time category	N (int)	N (cont)	Risk group	Assessor (IA or BICR)	Intervention median (95%CI)	Control median (95%CI)	HR (95%CI)
1L											
JAVELIN Renal 101	Haanen 2023	Ave+ axi	Suni	2-3yr	260	141	Overall	IA	19.4 (15.2, 22.3)	14.5 (8.8, 17.1)	
CROSS-J-RCC	Tomita (2020)	Suni	Sora	4-5yr	60	64	Overall	Unclear	32.0	14.9	
CheckMate 9ER	Company submission	Cabo +nivo	Suni	3-4yr	181	92	Overall	BICR	22.08 (17.97, 26.02)	16.07 (11.07, 19.35)	
CLEAR	Motzer 2023	Pem+ lenv	Suni	4-5yr	253	131	Overall	BICR	26.7 (22.8, 34.6)	14.7 (9.4, 18.2)	
CheckMate 214	Motzer 2022	Nivo+ipi	Suni	5yr+	550	546	Overall	BICR	Not reached (59.0, NE)	24.8 (19.7, 30.1)	0.49 (0.35, 0.68)
JAVELIN Renal 101	Haanen 2023	Ave+ axi	Suni	2-3yr	71	44	Fav	IA	22.6 (15.2, 31.7)	20.8 (14.5, 24.9)	
JAVELIN Renal 101	Haanen 2023	Ave+ axi	Suni	2-3yr	189	97	Int/poor	IA	19.3 (13.9, 22.1)	9.8 (7.0, 15.3)	
2L+											
AXIS	Rini 2011	Axi	Sora	1-2yr	361	362	Overall	NR	11 (7.4, NE)	10.6 (8.8, 11.5)	
NCT01136733	Motzer 2015	Lenv+ evero	Evero	1-2yr	51	50	Overall	NR	13 (3.7, NE)	8.5 (7.5, 9.4)	

Trial name	First author	Int. name	Control name	Follow-up time category	N (int)	N (cont)	Risk group	Assessor (IA or BICR)	Intervention median (95%CI)	Control median (95%CI)	HR (95%CI)
TIVO-3	Verzoni 2021	Tivo	Sora	1-2yr	175	175	Overall	IA	20.3 (9.8, 29.9)	9 (3.7, 16.6)	
CheckMate 025	Motzer 2020	Nivo	Evero	5yr+	410	411	Overall	NR	18.2 (12.9, 25.8)	14 (8.3, 19.2)	

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; BICR, blinded independent central review; CI; confidence interval; evero, everolimus; HR, hazard ratio; IA, investigator-assessed; ipi, ipilimumab; lenv, lenvatinib; nivo, nivolumab; NR, not reported; pazo, pazopanib; PBO, placebo; pem, pembrolizumab; RECIST, Response Evaluation Criteria in Solid Tumours sora, sorafenib; suni, sunitinib; tivo, tivozanib; yr, year

3.3.3.5. Time to next treatment

The time to next treatment was only available for two trials, CheckMate 9ER and CheckMate 214, with data provided by the manufacturers as part of this appraisal. Data was available in the 1st line setting for an overall risk, favourable risk, and intermediate/poor risk population from CheckMate 9ER and for an intermediate/poor risk population in CheckMate 214. CheckMate 9ER evaluated cabozantinib + nivolumab and CheckMate 214 evaluated nivolumab + ipilimumab, both trials involved a comparison with sunitinib. Median time to next treatment for sunitinib in CheckMate 9ER was [redacted] across all risk groups, whereas this was [redacted] the length of time for sunitinib in the intermediate/poor risk group in Checkmate 214. It is likely this was due to the non-standard definition used, which was the survival time from end of therapy in patients who never received subsequent systemic treatment, or the time from end of therapy until subsequent systemic treatment in patients who received subsequent systematic treatment. Both [redacted] and [redacted] showed a [redacted] time to next treatment than [redacted].

Table 17: Time to next treatment in prioritised included trials

Trial name	First author	Int. name	Con. name	Risk group	Line	Follow-up time category	N (int)	N (con)	Median (int)	95% CI (int)	Median (cont)	95% CI (con)	Prop (int)	Prop (con)
CheckMate 9ER	Company clarification response	Cabo+ nivo	Suni	Overall	1L	3-4yr	263	288						
CheckMate 9ER	Company clarification response	Cabo+ nivo	Suni	Fav	1L	3-4yr	60	57						
CheckMate 9ER	Company clarification response	Cabo+ nivo	Suni	Int / poor	1L	3-4yr	203	231						
CheckMate 214	Stakeholder submission	Nivo+ipi	Suni	Int / poor	1L	5yr+	423	416						

Abbreviations: cabo, cabozantinib; CI; confidence interval; con, control; int, intermediate / intervention; nivo, nivolumab; NE, non-evaluable; NR, not reported; suni, sunitinib; yr, year

3.3.3.6. Time on treatment

Time on treatment was available for eight trials evaluating 1st line treatment in the overall risk group: CLEAR, CROSS-J-RCC, SWITCH, SWITCH II, COMPARZ, CheckMate 9ER, CheckMate 214 and TIVO-1. For CheckMate 9ER, both the duration of treatment and the time to discontinuation was reported, whereas all other studies reported only the duration of treatment. Data are shown in Table 18.

The median duration of treatment was reported for sunitinib in six trials[91,92,100,106,109,110] in the overall risk group, which ranged between 6.7 and 10.1 months and in two trials in the intermediate/poor risk population, which ranged between 6.1 and 7.1 months. Median duration of response in the overall risk population was also available for pazopanib (2 trials[91,101]), cabozantinib + nivolumab (1 trial[110]), nivolumab + ipilimumab (1 trial[106]), pembrolizumab + lenvatinib (1 trial[109]) and tivozanib (1 trial[103]). In descending order, median treatment duration was 21.8 months for cabozantinib + nivolumab, 17 months for pembrolizumab + lenvatinib, 12 months for tivozanib, 7.9 for nivolumab + ipilimumab, 5.7 to 8 months for pazopanib. Treatment duration was often similar between trial arms,

though cabozantinib + nivolumab, pembrolizumab + lenvatinib and tivozanib were each associated with a clear longer treatment duration than their comparator.

Duration of response was only available from one trial in the favourable risk population. This data showed that duration of treatment was longer in both arms (cabozantinib + nivolumab and sunitinib) than in the overall risk population, though the increase for cabozantinib + nivolumab was negligible ([redacted] compared to 21.8 months). Sunitinib was associated with more than 4 months' additional treatment duration in the favourable risk population compared to the overall group.

Duration of treatment was reported in three trials in the intermediate/poor risk group. Treatment duration with sunitinib was similar across all three trials, ranging from 6.1 to 7.1, and was comparable with the overall risk population. Median duration of treatment for cabozantinib and for nivolumab + ipilimumab were no different than their comparator, sunitinib; 8.4 and [redacted], respectively. Treatment duration was substantially longer for cabozantinib + nivolumab than sunitinib, however, at a median of [redacted].

In the 2nd line-plus population, four trials reported duration of treatment, all in an overall risk population: RECORD-1, TIVO-3, AXIS and CheckMate 025. Evidence was available for everolimus (2 trials), nivolumab (1 trial), axitinib (1 trial), tivozanib (1 trial), sorafenib (2 trials), and placebo (1 trial).

In descending order, duration of treatment was a mean of 8.2 months for axitinib, median 6.4 months for tivozanib, a median of 4.6 to a mean of 5.2 months for sorafenib, a median of 4.6 months for everolimus in RECORD-1, and a median of 2.0 months for placebo.

Axitinib and tivozanib each showed a longer treatment duration than their comparator, sorafenib, and everolimus had a longer treatment duration than placebo.

Table 18: Time on treatment in prioritised included trials

Trial name	First author	Year	Int name	Control name	Follow-up time category	N (int)	N (cont)	Risk group	ToT (int)	ToT (control)
1L										
SWITCH	Eichelberg	2015	Sora	Suni	<1yr	177	176	Overall	Mean 8.7 months (SD 8.6)	Mean 10.1 months (SD 10.2)
COMPARZ	Motzer	2013	Pazo	Suni	1-2yr	557	553	Overall	Median 8 (range 0, 38)	Median 7.6 range 0, 38)
SWITCH II	Retz	2019	Sora	Pazo	3-4yr	189	188	Overall	Median 2.1 (range 0.3, 21.4)	Median 5.7 (range 0.3, 43.3)
CheckMate 9ER	Company submission	2023	Cabo+nivo	Suni	3-4yr	323	328	Overall	Median 21.8 (IQR 8.8, 34.0)	Median 8.9 (IQR 2.9, 20.7)
CheckMate 9ER	Company submission	2023	Cabo+nivo	Suni	3-4yr	323	328	Overall	[redacted]	[redacted]
CROSS-J-RCC	Tomita	2020	Suni	Sora	4-5yr	60	64	Overall	Median 6.7 (95%CI NR);	Median 5.9 (95%CI NR);
CheckMate 214	Motzer	2022	Nivo+ipi	Suni	5yr+	550	546	Overall	Median 7.9 (IQR 2.1, 21.8)	Median 7.8 (IQR 3.5, 19.6)
CLEAR	Motzer	2023	Pem+lenv	Suni	NR	355	357	Overall	Median 17 (95%CI 9.4, 25.4)	Median 7.8 (95%CI 3.7, 17.8)
TIVO-1	Motzer	2013	Tivo	Sora	NR	259	257	Overall	Median 12 (95%CI NR);	Median 9.5 (95%CI NR)
CheckMate 9ER	Company submission	2023	Cabo+nivo	Suni	3-4yr	74	71	Fav	[redacted]	[redacted]

Trial name	First author	Year	Int name	Control name	Follow-up time category	N (int)	N (cont)	Risk group	ToT (int)	ToT (control)
CABOSUN	Choueiri	2018	Cabo	Suni	2-3yr	78	72	Int/poor	Median 8.39 (95%CI 5.72, 8.39)	Median 7.09 (95%CI 5.09, 6.68)
CheckMate 9ER	Company submission	2023	Cabo+nivo	Suni	3-4yr	246	249	Int/poor	[redacted]	[redacted]
CheckMate 214	Stakeholder submission	2023	Nivo+ipi	Suni	5yr+	423	416	Int/poor	[redacted]	[redacted]
2L+										
RECORD-1	Motzer	2010	Evero	Placebo	1-2yr	277	139	Overall	Median 4.64 (95%CI NR);, range 0.62, 4.96)	Median 1.97 (95%CI NR); range 0.69, 6.4)
TIVO-3	Rini	2020	Tivo	Sora	1-2yr	175	175	Overall	Median 6.48 (95%CI NR); IQR 3.7, 14.0)	Median 4.64 (95%CI NR); IQR 2.3, 7.7)
AXIS	Motzer	2013	Axi	Sora	3-4yr	361	362	Overall	Mean 8.2 (SD NR, range <0.1, 33.4)	Mean 5.2 (SD NR, range 0.2, 34.1)

Abbreviations: axi, axitinib; cabo, cabozantinib; evero, everolimus; ipi, ipilimumab; IQR, interquartile range; lenv, lenvatinib; nivo, nivolumab; NR, not reported; pazo, pazopanib; pem, pembrolizumab; suni, sunitinib; tivo, tivozanib; ToT, time on treatment; yr, year

Note: *Time to discontinuation

3.3.3.7. Adverse events of treatment

Discontinuation due to adverse events

No studies reported separate adverse event rate data in population subgroups, and so all evidence was reported in an overall risk group or, in the case of one trial in the 1st line setting, in an intermediate/poor risk population that was the entire the trial sample.

1st line

In the 1st line setting, nine studies reported the rate of discontinuation due to adverse events in an overall risk population. All trials involved a comparison with sunitinib (7 trials) and/or sorafenib (4 trials). Pazopanib was evaluated in two trials, all other interventions (tivozanib, pembrolizumab + lenvatinib, nivolumab + ipilimumab, cabozantinib + nivolumab, and avelumab + axitinib) were evaluated in only one trial.

The rate of discontinuation due to AEs ranged between 11.5% to 28.4% for sunitinib and 7.0% to 32.3% for sorafenib, with no clear relationship with the length of follow-up. Avelumab + axitinib, cabozantinib + nivolumab, nivolumab + ipilimumab, and pembrolizumab + lenvatinib all had a higher rate of discontinuation due to adverse events than sunitinib. Rates of discontinuation were particularly high for avelumab + axitinib, cabozantinib + nivolumab, pembrolizumab + lenvatinib and nivolumab + ipilimumab, where the rate of discontinuation exceeded 30% of the trial arm. Rates of discontinuation for tivozanib were comparable with sunitinib, while rates of discontinuation for pazopanib were comparable with sunitinib and lower than sorafenib.

One trial reported discontinuation due to adverse events in an intermediate/poor risk population. The rate of discontinuation was similar for cabozantinib and sunitinib.

2nd line-plus

Seven trials reported the rate of discontinuation due to adverse events in the 2nd line-plus setting. Of these, five trials evaluated everolimus, two trials evaluated sorafenib, two trials evaluated axitinib, and the remaining treatments (cabozantinib, everolimus + lenvatinib, tivozanib, and nivolumab) were each evaluated in one trial. Rates of discontinuation due to adverse events ranged between 0% and 16.1% for everolimus, 12.4% to 29.7% for sorafenib, and 0% to 7.5% for axitinib. Rates of discontinuation were generally lower than in the 1st line setting, and relative effects were therefore imprecise. There was a trend for a higher rate of discontinuation following everolimus + lenvatinib than everolimus alone, otherwise rates of discontinuation were similar between everolimus and cabozantinib, nivolumab, and axitinib. With the exception of TIVO-1, where rates of discontinuation appeared higher than other trials, rates of discontinuation were generally less than 15% of the trial arm.

Table 19: Discontinuation due to adverse events in prioritised included trials

Trial name	Author (year)	Int name	Control name	Follow-up time category	N (int)	N (cont)	Risk group	% (int)	% (control)	OR (95%CI)
1L										
SWITCH	Eichelberg (2015)	Sora	Suni	<1yr	182	183	Overall	18.13%	28.42%	0.56 (0.34, 0.92)
CLEAR	Motzer (2021)	Pem+lenv	Suni	1-2yr	355	357	Overall	16.90%	11.48%	1.57 (1.02, 2.40)
COMPARZ	Motzer (2013)	Pazo	Suni	1-2yr	557	553	Overall	24.24%	20.25%	1.26 (0.95, 1.67)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	2-3yr	442	444	Overall	31.22%	15.99%	2.38 (1.73, 3.30)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	2-3yr	323	328	Overall	36.84%	20.43%	2.27 (1.60, 3.23)
SWITCH II	Retz (2019)	Sora	Pazo	3-4yr	189	188	Overall	32.28%	23.40%	1.56 (0.99, 2.46)
SWOG 1500	Pal (2021)	Cabo	Suni	3-4yr	44	46	Overall	22.73%	23.91%	0.94 (0.35, 2.49)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	4-5yr	60	64	Overall	21.67%	18.75%	1.20 (0.50, 2.88)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	5yr+	550	546	Overall	34.18%	19.41%	2.16 (1.64, 2.84)
TIVO-1	Motzer (2013)	Tivo	Sora	NR	260	257	Overall	7.31%	7.00%	1.05 (0.54, 2.04)
CABOSUN	Choueiri (2018)	Cabo	Suni	NR	79	78	Int/poor	20.25%	20.51%	0.98 (0.45, 2.14)
2L+										
METEOR	Choueiri (2016)	Cabo	Evero	1-2yr	330	328	Overall	12.12%	10.37%	1.19 (0.73, 1.94)
NCT01136733	Motzer (2015)	Lenv+evero	Evero	1-2yr	51	50	Overall	17.65%	10.00%	1.93 (0.60, 6.22)

Trial name	Author (year)	Int name	Control name	Follow-up time category	N (int)	N (cont)	Risk group	% (int)	% (control)	OR (95%CI)
RECORD-1	Motzer (2010)	Evero	Placebo	1-2yr	277	139	Overall	13.00%	1.44%	10.23 (2.43, 43.16)
TIVO-3	Zengin (2020)	Tivo	Sora	1-2yr	175	175	Overall	20.57%	29.71%	0.61 (0.38, 1.00)
AXIS	Motzer (2013)	Axi	Sora	3-4yr	361	362	Overall	7.48%	12.43%	0.57 (0.34, 0.94)
CheckMate 025	Motzer (2020)	Nivo	Evero	5yr+	410	411	Overall	13.90%	16.06%	0.84 (0.57, 1.24)
BERAT	Grunwald (2022)	Evero	Axi	NR ('short')	5	5	Overall	0.00%	0.00%	-

Abbreviations axi, axitinib; cabo, cabozantinib; con, control; evero, everolimus; int, intervention; ipi, ipilimumab; IQR, interquartile range; lenv, lenvatinib; nivo, nivolumab; NR, not reported; pazo, pazopanib; pem, pembrolizumab; suni, sunitinib; tivo, tivozanib; ToT, time on treatment; yr, year

Note: *Time to discontinuation

Grade 3+ adverse events

1st line

Nine trials reported the rate of Grade 3+ AEs in an overall risk population in the 1st line setting. All trials involved a comparison with sunitinib (7 trials) and/or sorafenib (4 trials). Pazopanib was evaluated in 2 trials, all other treatments (pembrolizumab + lenvatinib, avelumab + axitinib, cabozantinib + nivolumab, nivolumab + ipilimumab, and tivozanib) were each evaluated in one trial. All interventions were associated with high rates of grade 3+ events. Rates ranged between 64.5% to 83.3% for sunitinib, 57.1% to 75.0% for sorafenib, and 62.2% to 74.0% for pazopanib. Rates for all other treatments exceeded 60% of the trial arm and were particularly high (exceeding three quarters of the sample) following cabozantinib + nivolumab, pembrolizumab + lenvatinib, and avelumab + axitinib. The risk of grade 3+ AEs was lower for tivozanib than sorafenib, and for nivolumab + ipilimumab than sunitinib, each evaluated in one trial.

In an intermediate/poor risk population, there was a small increased risk of grade 3+ AEs following cabozantinib in comparison with sunitinib, but the difference was not statistically significant. In general, rates of grade 3+ events were comparable with those reported in the 1st line setting.

2nd line-plus

Four trials reported rates of grade 3+ adverse events in the 2nd line setting, all in an overall risk population. All trials involved a comparison with everolimus, while the other treatments (cabozantinib, everolimus + lenvatinib, nivolumab, and axitinib) were all evaluated in one trial. There was wide variation in the rates of grade 3+ adverse events across trials, with rates for everolimus ranging between 36.8% (in the trial with the longest follow-up) to 58.8%. The highest risk was reported for axitinib, where 80% of participants experienced a Grade 3+ AE. Risk was also high for cabozantinib and everolimus + lenvatinib, where more than 70% of participants experienced a Grade 3+ event. Axitinib, cabozantinib, and everolimus + lenvatinib were each associated with an increased risk of Grade 3+ events relative to everolimus, while nivolumab had a lower risk of events relative to everolimus.

Table 20: Grade 3+ adverse events in prioritised included trials

Trial name	Author (year)	Intervention name	Control name	Follow-up time category	N (int)	N (con)	Risk group	% (int)	% (con)	OR (95% CI)
1L										
SWITCH	Eichelberg (2015)	Sora	Suni	<1yr	182	183	Overall	64.29 %	64.48 %	0.99 (0.65, 1.52)
CLEAR	Motzer (2021)	Pem+lenv	Suni	1-2yr	355	357	Overall	81.69 %	68.35 %	2.07 (1.46, 2.93)
COMPARZ	Motzer (2013)	Pazo	Suni	1-2yr	557	553	Overall	73.97 %	72.69 %	1.07 (0.82, 1.39)
JAVELIN Renal 101	Haanen (2023)	Ave+axi	Suni	2-3yr	442	444	Overall	79.64 %	76.58 %	1.20 (0.87, 1.65)
SWITCH II	Retz (2019)	Sora	Pazo	3-4yr	189	188	Overall	57.14 %	62.23 %	0.81 (0.54, 1.22)

Trial name	Author (year)	Intervention name	Control name	Follow-up time category	N (int)	N (con)	Risk group	% (int)	% (con)	OR (95% CI)
SWOG 1500	Pal (2021)	Cabo	Suni	3-4yr	44	46	Overall	72.73 %	67.39 %	1.29 (0.52, 3.19)
CheckMate 9ER	Company submission (2023)	Cabo+nivo	Suni	3-4yr	323	328	Overall	[redacted]	[redacted]	1.70 (1.16, 2.49)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	4-5yr	60	64	Overall	83.33 %	75.00 %	1.67 (0.69, 4.03)
CheckMate 214	Motzer (2022)	Nivo+ipi	Suni	5yr+	550	546	Overall	67.82 %	76.23 %	0.66 (0.50, 0.86)
TIVO-1	Motzer (2013)	Tivo	Sora	NR	260	257	Overall	61.15 %	69.65 %	0.69 (0.48, 0.99)
CABOSUN	Choueiri (2018)	Cabo	Suni	NR	79	78	Int/poor	67.09 %	60.26 %	1.34 (0.70, 2.58)
2L+										
METEOR	Choueiri (2016)	Cabo	Evero	1-2yr	330	328	Overall	71.21 %	58.84 %	1.73 (1.25, 2.39)
NCT01136733	Motzer (2015)	Lenv+evero	Evero	1-2yr	51	50	Overall	70.59 %	50.00 %	2.40 (1.06, 5.44)
CheckMate 025	Motzer (2020)	Nivo	Evero	5yr+	410	411	Overall	21.40 %	36.80 %	0.47 (0.34, 0.64)
BERAT	Grunwald (2022)	Evero	Axi	NR ('short')	5	5	Overall	40.00 %	80.00 %	0.17 (0.01, 2.82)

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; con, control; evero, everolimus; int, intervention; ipi, ipilimumab; lenv, lenvatinib; nivo, nivolumab; pazo, pazopanib; OR, odds ratio; pem, pembrolizumab; suni, sunitinib; tivo, tivozanib; yr, year

3.3.3.8. Health-related quality of life

1st line

Overall risk

Six trials reported HRQoL in an overall risk population in the 1st line: all six trials reported a disease specific HRQoL (FKSI total [4 trials] and FKSI DRS [2 trials]) and four trials reported generic HRQoL (EQ-5D index [3 trials] and EQ-5D VAS [1 trial]). This section focusses condition specific analysis on the FKSI total as the more comprehensive and frequently reported scale. All trials involved a comparison with sunitinib (four trials) or sorafenib (two trials). One trial was a mix of 1st and 2nd line (TIVO-1).

Baseline FKSI total scores were reported to be between 58.4 – 60.1 (reported in 2 trials; CheckMate 9ER and CheckMate 214) and baseline FKSI DRS scores were 29.2 – 31.3 [CLEAR and TIVO-1]. Baseline EQ-5D scores ranged between 0.73 – 0.83 (CheckMate 9ER, CLEAR, TIVO-1). None of the trials reported meaningful differences in HRQoL between treatment arms according to established MID thresholds.[115-118] Four trials reported mean change in HRQoL in each arm (CLEAR, SWITCH II, CheckMate 214 and TIVO-1), which showed that pembrolizumab + lenvatinib, sunitinib, sorafenib and pazopanib were all associated with meaningful reductions in disease-specific HRQoL over time, whereas there was no change for nivolumab and ipilimumab. There were reductions in generic HRQoL following pembrolizumab + lenvatinib, sunitinib, tivozanib and sorafenib, but these were not greater than the threshold for a minimally important difference.

Favourable risk

Two trials reported HRQoL in a favourable risk population in the 1st line: one trial reported both disease-specific and generic HRQoL (FKSI-DRS and EQ-5D index) and one trial reported only disease-specific HRQoL (FKSI total). Neither trial reported baseline HRQoL. The CLEAR trial reported a bigger reduction in FKSI-DRS scores within the year following treatment with pembrolizumab + lenvatinib than sunitinib, and this approached the threshold for a minimally important difference. Both arms experienced meaningful reductions in both disease-specific and generic HRQoL during this time, which passed or approached the threshold for a minimally

important difference. Arm-specific changes in HRQoL were not reported for C CheckMate 9ER, but there was no meaningful difference in FKSI total scores between cabozantinib + nivolumab and sunitinib.

Intermediate/poor risk

Three trials reported HRQoL in an intermediate/poor risk population in the 1st line: three trials reported disease-specific HRQoL (FKSI total [2 trials] and FKSI-DRS [1 trial]) and two trials reported generic HRQoL (EQ-5D index [1 trial] and EQ-5D VAS [1 trial]). All trials involved a comparison with sunitinib. Treatment with sunitinib was followed by meaningful reductions in HRQoL [2 trials]. Pembrolizumab + lenvatinib was associated with a smaller reduction in disease-specific and generic HRQoL [1 trial], while there was no meaningful change in disease-specific HRQoL following nivolumab and ipilimumab. Cabozantinib + nivolumab showed a meaningful benefit for HRQoL over sunitinib, but baseline scores and the change in HRQoL in each arm was not provided. Numerical benefits were also shown for nivolumab + ipilimumab and pembrolizumab + lenvatinib as compared to sunitinib.

2nd line-plus

Four trials reported HRQoL in the 2nd line-plus, all in an overall risk population: four trials reported disease-specific HRQoL (FKSI total [3 trials] and FKSI-DRS 9 [1 trial]). Three trials involved a comparison with everolimus (vs. cabozantinib, sorafenib and nivolumab) and one trial was a comparison with sorafenib (vs axitinib). HRQoL increased in both arms of the BERAT trial (everolimus vs axitinib), but otherwise HRQoL in the trials remained the same or decreased following treatment. There was a difference in disease-specific HRQoL between nivolumab and everolimus, with higher HRQoL at follow up for those receiving nivolumab, but arm-specific change in HRQoL was not reported, and there was no difference in generic HRQoL between arms. There was no difference in disease-specific HRQoL between cabozantinib and everolimus.

Table 21: HRQoL data in prioritised included trials

Trial name	First author	Int name	Con name	Risk gp	Definition of event and censor variables	Measure	Follow-up time category	N (int)	N (con)	BL (int)	BL (con)	Outcome (int)	Outcome (con)	Mean diff (95%CI)
1L														

Trial name	First author	Int name	Con name	Risk gp	Definition of event and censor variables	Measure	Follow-up time category	N (int)	N (con)	BL (int)	BL (con)	Outcome (int)	Outcome (con)	Mean diff (95%CI)
CLEAR	Motzer (2022)	Pem + lenv	Suni	All	Disease specific HRQoL	FKSI-DRS. Mean change, LS mean difference	<1yr	355	357	31.28 (4.41)	30.89 (4.90)	Mean: -1.75 (SE 0.59)	Mean: -2.19 (SE 0.66)	0.44 (-1.11, 2.00)
COMPARZ	Motzer (2013)	Pazo	Suni	All	Disease specific HRQoL	FKSI total score. Difference in mean change score intervention vs control	1-2yr	377	408	NR	NR	NR	NR	1.41 (NR)
CheckMate 9ER	Cella (2022)	Cabo + Nivo	Suni	All	Disease specific HRQoL	FKSI total, LS mean change score. HR is time to deterioration	1-2yr	323	328	58.74 (10.57)	58.39 (9.92)	NR	NR	2.38 (1.20, 3.56)
SWITCH II	Retz (2019)	Sora	Pazo	All	Disease specific HRQoL	FKSI-10	3-4yr	183	183	NR	NR	Mean: -3.1 (SD NR)	Mean: -3.7 (SD NR)	NR
CheckMate 214	Motzer (2022)	Nivo + Ipi	Suni	All	Disease specific HRQoL	FKSI-19 LS mean change	5yr+	550	546	60.1	59.1	Mean: 0.36 (SD NR)	Mean: -1.51 (SD NR)	1.87 (0.95, 2.79)
CLEAR	Motzer (2022)	Pem + lenv	Suni	All	Generic HRQoL	EQ5D-Index, Mean change, LS mean difference	<1yr	355	357	0.83 (0.19)	0.81 (0.22)	Mean: -4 (SE 0.9)	Mean: -6 (SE 1.1)	2 (0, 5)
CheckMate 9ER	Cella (2022)	Cabo	Suni	All	Generic HRQoL	EQ-5D-3L UK index score, LS mean change score. HR is the time to deterioration	1-2yr	323	328	0.78 (0.25)	0.73 (0.29)	NR	NR	0.04 (0.01, 0.07)
CheckMate 214	Cella (2020)	Nivo + ipi	Suni	All	Generic HRQoL	EQ-5D VAS LS mean using MMRM	5yr+	550	546	NR	NR	NR	NR	2.4 (0.4, 4.5)
TIVO-1	Motzer (2013)	Tivo	Sora	All	Disease specific HRQoL	FKSI-DRS LS mean change from baseline	NR	256	250	29.16 (4.77)	29.35 (5.10)	Mean: -0.94 (SE 0.33)	Mean: -0.93 (SE 0.34)	NR
TIVO-1	Motzer (2013)	Tivo	Sora	All	Generic HRQoL	EQ-5D. This is a LS mean change score from baseline	NR	256	250	0.73 (0.25)	0.73 (0.26)	Mean: -0.05 (SD 0.02)	Mean: -0.06 (SD 0.02)	NR
CLEAR	Motzer (2022)	Pem + lenv	Suni	Fav	Disease specific HRQoL	FKSI-DRS. Mean change, LS mean difference	<1yr	110	124	NR	NR	Mean: -4.67 (SE 0.96)	Mean: -3.69 (SE 0.98)	-0.97 (-3.58, 1.61)
CheckMate 9ER	Cella (2023)	Cabo + nivo	Suni	Fav	Disease specific HRQoL	FKSI total, LS mean change score	1-2yr	74	72	NR	NR	NR	NR	-0.44 (-2.63, 1.75)
CLEAR	Motzer (2022)	Pem + lenv	Suni	Fav	Generic HRQoL	EQ5D-Index, Mean change, LS mean difference	<1yr	110	124	NR	NR	Mean: -8 (SE 1.4)	Mean: -6 (SE 1.5)	-2 (-6, 2)

Trial name	First author	Int name	Con name	Risk gp	Definition of event and censor variables	Measure	Follow-up time category	N (int)	N (con)	BL (int)	BL (con)	Outcome (int)	Outcome (con)	Mean diff (95%CI)
CLEAR	Motzer (2022)	Pem + lenv	Suni	Int/poor	Disease specific HRQoL	FKSI-DRS. Mean change, LS mean difference	<1yr	243	229	NR	NR	Mean: -0.72 (SE 0.86)	Mean: -1.42 (SE 0.96)	0.67 (-1.25, 2.58)
CheckMate 9ER	Cella (2023)	Cabo + nivo	Suni	Int/poor	Disease specific HRQoL	FKSI total, LS mean change score. HR is time to deterioration	1-2yr	249	256	NR	NR	NR	NR	3.33 (1.96, 4.70)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	Int/poor	Disease specific HRQoL	FKSI-19 LS mean change	5yr+	425	422	NR	NR	Mean: 0.9 (SD NR)	Mean: -1.75 (SD NR)	2.65 (1.60, 3.70)
CLEAR	Motzer (2022)	Pem + lenv	Suni	Int/poor	Generic HRQoL	EQ5D-Index, Mean change, LS mean difference	<1yr	243	229	NR	NR	Mean: -3 (SE 1.5)	Mean: -7 (SE 1.7)	5 (1, 8)
CheckMate 214	Cella (2020)	Nivo + ipi	Suni	Int/poor	Generic HRQoL	EQ-5D VAS LS mean using MMRM	5yr+	425	422	NR	NR	NR	NR	3.3 (1.0, 5.6)
2L+														
METEOR	Cella (2018)	Cabo	Evero	All	Disease specific HRQoL	FKSI-19 LS mean change	<1yr	324	313	NR	NR	Mean: -3.483 (SD NR)	Mean: -2.214 (SD NR)	-1.269 (-1.864, -0.675)
AXIS	Motzer (2013)	Axi	Sora	All	Disease specific HRQoL	FKSI-15	1-2yr	NR	NR	43.2 (8.4)	43.3 (8.2	Mean: 38.9 (SD 9.5)	Mean: 39.1 (SD 8.9)	NR
CheckMate 025	Cella (2016)	Nivo	Evero	All	Disease specific HRQoL	FKSI-DRS mean change	1-2yr	361	343	30.2 (4.4)	30.8 (4.8)	NR	NR	1.6 (1.4, 1.9)
BERAT	Grunwald (2022)	Evero	Axi	All	Disease specific HRQoL	FKSI-10		2	1	16.25 (SD 5.0)	19.7 (SD 2.89)	Mean: 22 (SD 1.41)	Mean: 15 (SD NR)	NR
METEOR	Cella (2018)	Cabo	Evero	All	Generic HRQoL	EQ-5D Index LS mean change	<1yr	323	314	NR	NR	Mean: -0.02 (SD NR)	Mean: -0.02 (SD NR)	-0.002 (-0.018, 0.014)
CheckMate 025	Cella (2016)	Nivo	Evero	All	Generic HRQoL	EQ-5D mean change	1-2yr	361	344	0.78 (0.24)	0.78 (0.21)	NR	NR	0.04 (0.02, 0.07)
AXIS	Cella (2013)	Axi	Sora	All	Generic HRQoL	EQ-5D estimated using repeated measures analysis adjusting for time	NR	NR	NR	NR	NR	Mean: 0.71 (SD NR)	Mean: 0.69 (SD NR)	NR

Abbreviations: axi, axitinib; BL, baseline; cabo, cabozantinib; con, control; evero, everolimus; HRQoL, health-related quality of life; int, intervention; ipi, ipilimumab; IQR, interquartile range; lenv, lenvatinib; nivo, nivolumab; NR, not reported; pazo, pazopanib; pem, pembrolizumab; suni, sunitinib; tivo, tivozanib;

3.4. Description and critique of the evidence presented by the company

The company submission for cabozantinib with nivolumab comprised a main submission, an appendix and a subsequent submission with updated efficacy data from CheckMate 9ER. The company also conducted a SLR to identify evidence relevant to the evaluation of cabozantinib with nivolumab. The company reported the synthesis of the identified evidence in a separate report, the findings of which we do not summarise in detail but contrast with our own network meta-analyses in Section 3.7.6. The EAG requested IPD from the company to enable the network meta-analysis and survival analysis to be run as robustly as possible, but this was not received.

3.4.1. Company's definition of the decision problem

The company's approach to the decision problem is presented in Table 22. The EAG broadly agreed with most decisions taken by the company, but disagreed on the full range of appropriate comparators, the relevance of time to next treatment, and the importance of risk group-specific analyses.

Table 22: Decision problem submitted by the company

	Final scope issued by NICE	Decision problem addressed in the company submission	Company rationale if different from the final NICE scope	EAG response
Population	Patients with untreated advanced or metastatic renal cell carcinoma	Patients with untreated advanced or metastatic renal cell carcinoma	NA	The EAG agrees the scope has been fulfilled.
Intervention	Cabo+nivo as a 1L therapy in untreated advanced or metastatic renal cell carcinoma.	Cabo+nivo as a 1L therapy in untreated advanced or metastatic renal cell carcinoma.	NA	The EAG agrees the scope has been fulfilled.
Comparator(s)	• Pazo • Tivo • Suni • Cabo (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Nivo+ipi (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Pem+lenv (only for intermediate- or poor-risk disease as defined in the IMDC criteria)	• Pazo • Suni • Cabo (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Nivo+ipi (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Pem+lenv (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Ave+axi	Although currently in the CDF, ave+axi is available to an all-risk aRCC NHS England population. Significantly, ave+axi has been in the CDF for over four years now, an unusual length of time for the CDF. Additionally, as highlighted by a recent ABPI report, the majority of therapies (78%) exit the CDF into routine commissioning suggesting that ave+axi is also expected to enter routine commissioning. Therefore, ave+axi should be considered as a relevant comparator by NICE and is discussed as such in our submission. Tivo is not included as a comparator in this submission as the NMA that was conducted to support Ipsen HTA submissions for other countries determined tivo was not widely used in practice. There are data	The EAG disagrees that ave+axi is a relevant comparator for this appraisal. This is because it is not expected that axi with ave will exit managed access by the time the Committee discusses this specific access decision, as is consistent with the standard NICE position. While the EAG identified and synthesised clinical evidence relating to this drug combination, it stresses that this is for validation only and not on the basis of a comparison with routinely commissioned treatment. The EAG also disagrees that tivo should be excluded. While it could not readily be included in evidence networks for OS, it nevertheless is an important drug to be considered in this analysis. This was discussed in response to clarification question A1, where the company cited

	Final scope issued by NICE	Decision problem addressed in the company submission	Company rationale if different from the final NICE scope	EAG response
	• Active surveillance		available to link and create a network. However, tivo has been assessed as an equivalent treatment to suni and pazo in previous NICE submissions. Active surveillance is not included in this submission; as discussed in the scoping call on 16th January 2023, active surveillance is usually used in 1L favourable risk patients and involves a wait-period before therapy is administered. Therefore, it is not relevant to this submission.	market share data to justify its exclusion. The EAG did not agree that this was an appropriate rationale. The UK RWE sourced by the EAG indicated tivo and cabo have a similar market share at 1L. Finally, the EAG agrees that active surveillance is not a relevant comparator in this appraisal.
Outcomes	• Overall survival • PFS • Response rates • DoR • Time on treatment/Time to next treatment • Adverse effects of treatment • HRQoL	• Overall survival • PFS • Response rates • DoR • Time on treatment • Adverse effects of treatment • HRQoL	Time to next treatment is not presented in this submission as it is not of relevance to the decision problem.	The EAG disagrees that time to next treatment is irrelevant. The receipt of subsequent lines of treatment is an important clinical outcome. This endpoint was considered relevant by the clinicians at the scoping workshop.
Groups to be considered	Intermediate-/poor-risk advanced metastatic RCC as defined in the IMDC criteria	Patients with untreated advanced or metastatic renal cell carcinoma	Cabo+nivo is indicated for an all-risk population of 'patients with untreated advanced or metastatic renal cell carcinoma' and should be appraised in line with this indication [69,70]. The phase 3 CheckMate 9ER trial of cabo+nivo compared to suni demonstrated consistent clinical	The EAG disagrees with this assertion. Risk group is known to be an important prognostic factor as well as an important effect modifier across a range of RCC treatments. As a result, subgroup-specific evidence is highly probative. We also note that risk-based subgroups were

	Final scope issued by NICE	Decision problem addressed in the company submission	Company rationale if different from the final NICE scope	EAG response
			benefits across all patients, irrespective of prognostic risk profile.	considered in the previous MTA.[38]

Source: Company submission document A, table 1

Key: ABPI, The Association of the British Pharmaceutical Industry; aRCC, advanced renal cell carcinoma; CDF, Cancer Drugs Fund; DoR, duration of response; HRQoL, Health Related Quality of Life; IMDC, International Metastatic Renal Cell Carcinoma Database Consortium; MTA, multiple technology assessment; RCC, renal cell carcinoma; NA, not applicable; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NMA, network meta-analysis; PFS, Progression Free Survival; TTD, time to discontinuation

3.4.2. Literature review methods used by the company

Ipsen carried out its original SLR in 2017, which they subsequently updated in 2018. A new search was then designed and conducted on 5th June 2020, and further updated on 29th October 2021. The latest update of the search—presented in the CS—was performed on 2nd December 2022.

Ipsen's review was developed to support indirect treatment comparisons against cabozantinib + nivolumab. Because the review was developed for a range of markets, including the UK, their analysis ultimately focused on the following comparators:

- Nivolumab + Ipilimumab
- Avelumab + axitinib
- Pembrolizumab + axitinib
- Pembrolizumab + lenvatinib
- Cabozantinib monotherapy
- Sunitinib monotherapy
- Pazopanib monotherapy

However, a wider range of comparators than the above list was included in the search. The interventions searched for by Ipsen broadly overlapped with the interventions included in the PenTAG search, except that Ipsen did not include everolimus (which was included in the PenTAG search), while they did include temsirolimus, bevacizumab, interferon-alpha, and sorafenib (which were not included in the PenTAG search). Tivozanib was included within the company search terms but was then not considered within the evidence reviews as the company considered that it was not widely used.

Ipsen carried out literature searches for clinical evidence in OVID MEDLINE (including Epub Ahead of Print, In-Process & Other Non-Indexed Citations), Embase, the Cochrane Central Register of Controlled Trials via the Cochrane library, and the Health Technology Appraisal (HTA) Database (Ipsen Submission, Appendix C). They also performed a rapid appraisal search in the Cochrane library to identify existing systematic reviews in the topic area. The search strategies combine free-text and index terms for relevant cancers with free-text and index terms for relevant interventions. The Cochrane randomized controlled trial publication filter was used to limit the search results to RCTs (in MEDLINE and Embase). No language limits were applied.

Finally, Ipsen searched grey literature resources, including the trials registry ClinicalTrials.gov, online conference proceedings (searched only in updated search of October 2021), and the websites of national guideline and regulatory agencies, including NICE, Institute for Clinical and Economic Review (ICER), Haute Autorité de Santé (HAS), Gemeinsamer Bundesausschuss (G-BA), Canadian Agency For Drugs And Technologies In Health (CADTH) and International Network of Agencies for Health Technology Assessment (INAHTA), to identify European public assessment report (EPAR) and HTA documents (reviewed in the original search only).

In summary, Ipsen's literature searches use an appropriate range of databases and grey literature resources for the topic. The choice of free-text and index terms is also appropriate, and the searches have a reasonable balance of sensitivity and specificity.

The main difference between the reviews is the approach PenTAG took of first identifying systematic reviews and meta-analyses of the decision problem, followed by an update search since the search data of the most recent high quality systematic reviews. Ipsen, on the other hand, were updating their own initial 2017 literature review.

There are other small differences between the Ipsen and PenTAG searches. PenTAG only searched for HTA documents on the NICE website and the INAHTA database, while Ipsen also included ICER, HAS, G-BA, and CADTH (as their review is to be used to support submission across a wide range of markets). While in terms of clinical registries, Ipsen searched ClinicalTrials.gov only, while PenTAG searched ClinicalTrials.gov and the International Clinical Trials Registry Platform (ICTRP).

The search and review by Ipsen resulted in 142 reports being included in their NMA. When compared to the results of the PenTAG search, 76 of the 142 reports were retrieved by the PenTAG search, 49 were out of scope of the current decision problem, leaving 17 records in scope but not retrieved by the PenTAG search. These 17 records were appraised, and all added to the PenTAG review. Of the 17 records, 13 were conference abstracts (published from before the PenTAG RCT search), one was an FDA update in the Oncology Times, and three were full text journal articles (again from before the RCT search).

Only one of these 17 records contained new data not identified within the PenTAG search—a 2014 letter by Motzer et al[119] that contained the final overall survival outcomes of the COMPARZ trial.[91] This letter was included in the NMA of TA858 and was therefore also identified in citation chasing post the preliminary assessment report[38].

3.4.3. Analyses conducted by the company

CheckMate 9ER was a single-blind parallel group, RCT of cabozantinib + nivolumab comparing cabozantinib + nivolumab (n=323) against sunitinib (n=328). The trial included patients were those with advanced or metastatic RCC with a clear cell component (including patients with sarcomatoid features) who had also not received any prior systemic therapy. Patients could receive one prior adjuvant or neoadjuvant therapy if cancer recurrence was at least six months after the last dose (as is common across modern RCC trials) although only five patients did as use of adjuvant therapy was not common during the time of enrolment (Sept 2017 – May 2019). Though patients were required to have a Karnofsky performance score of at least 70%, all IMDC risk categories were included. Patients with active CNS metastases; active, known or suspected autoimmune disease; or with a range of comorbidities were excluded. CheckMate 9ER was conducted internationally across the USA, Europe and the Rest of the World with 21 patients enrolled from the UK.

A number of interim analyses were undertaken. In the company's original submission, the third database lock (median follow-up 32.9 months) was presented. This was later superseded by a fourth database lock with median follow-up of 44.0 months (minimum 36.5 for OS and PFS), which is the focus of discussion. The EAG regarded that controls for multiple analysis and multiple testing, including use of a hierarchical testing procedure, were appropriate. The EAG also regarded that assumptions underpinning sample size were, in some cases, unjustified (clarification response A7) but were not unreasonable given expected and observed trial results.

The primary outcome was PFS assessed via BICR according to FDA censoring rules. Analysis of the trial used standard methods. Differences between groups in survival outcomes used log-rank tests stratified by randomisation factors (IMDC category, PD-L1 tumour expression, and location of screening). Survival outcomes were further analysed using Cox proportional hazards models. In response to clarification question A21 on the validity of the proportional hazards assumption, the company provided results from tests on scaled Schoenfeld residuals and a check based on a visual examination of the log cumulative hazard plot. This was provided for OS and PFS outcomes in the ITT, intermediate/poor risk and favourable risk groups. The company argued based on these results that the assumption was met for all outcomes and groups except for OS in the favourable risk group. The EAG, however, believed that these assumptions were more tenuous than the company asserted; in the all-risk group, *p*-values from the tests of scaled Schoenfeld residuals were <0.10 for both outcomes, and it was not obvious

from any of the presented log-cumulative hazard plots that curves were indeed equidistant over the time horizon.

The EAG conducted quality assessment for all key trials, including CheckMate 9ER. This is presented in Section 3.3.2.5. The pivotal CheckMate 9ER trial was judged to have a high overall risk of bias because of a high risk of attrition bias (very high, differential overall attrition as well as dropouts due to discontinuation and disease progression, with reporting of single imputation of approaches to account for missing data). Random sequence generation was poorly reported, but pragmatically accepted as presenting low risk of bias due to the use of IxRS for randomisation. The EAG did not identify any specific additional conceptual concerns relating to the 44-month follow-up time point. However, the EAG noted that the company's explanation of the changes they made when they revised their data (clarification response A8) did not seem to encompass all of the changes made with minor differences observed for additional variables which were not noted as having been updated such as adverse events data. This creates some uncertainty related to data quality and consistency of definitions and datacuts.

The EAG noted several points in the outcome and design pattern of CheckMate 9ER that raise questions about the generalisability of this trial. Emerging observational evidence on the use of cabozantinib + nivolumab suggests that adverse event rates are possibly lower in routine practice than in the trial, with possible implications for observed effectiveness and relative dose intensity (clarification response A3). In addition, CheckMate 9ER enrolled a low number of UK patients (3.2%), which may indicate that effectiveness observed in the trial may not be reliably replicated in a UK treatment context (clarification response A5). CheckMate 9ER also included very few patients who had received a prior adjuvant treatment (n=5) due to the time period in which the trial was conducted, this does not align well with current and expected future practice in the UK following the recommendation of pembrolizumab in the adjuvant setting which impacts both on generalisability and on the achievability of the observed effect sizes. Finally, in response to clarification question A13, the company noted that [redacted] of patients receiving cabozantinib + nivolumab and [redacted] of patients receiving sunitinib continued to receive treatment post-progression, with mean duration of treatment beyond progression of [redacted] days and [redacted] days respectively. This is surprising given clinical advice that treatment generally ends at point of progression, and thus the trial may not reflect treatment patterns in the UK.

3.4.4. Results presented by the company

The EAG considered the most recent available data for each outcome to take precedence and therefore the focus of this section is the 44-month follow-up data, for which results are tabulated below (Table 23).

Table 23: Key results from 44-month follow up for CheckMate 9ER

Outcome	Cabo+nivo (n=323)	Suni (n=328)
BICR-observed PFS events	230	248
Median PFS months (95% CI)	16.56 (12.75, 19.48)	8.38 (6.97, 9.69)
Hazard ratio PFS (95% CI)	0.59 (0.49, 0.71), p<0.0001	
Median OS months (95% CI)	49.48 (40.31, N.E.)	35.52 (29.24, 42.25)
Hazard ratio OS (95% CI)	0.70 (0.56, 0.87)	
Increase in ORR (95% CI)	56.0% (50.4, 61.5)	28.0% (23.3, 33.2)
Median TTR months	2.83	4.32
Median DoR months	22.08 (17.97, 26.02)	16.07 (11.07, 19.35)
Median PFS-2 months	44.65 (35.94, N.A.)	25.07 (20.96, 32.36)
HR PFS-2 (95% CI)	0.63 (0.51, 0.78), p<0.0001	
Number of patients remaining on treatment[120]	57	32
Median TTD months	[redacted]	[redacted]
Number discontinued treatment	263 (82.2%)	288 (90.0%)
Proportion of discontinuers receiving a subsequent treatment	116/263 (44.1%)	148/288 (51.4%)
Most common type of subsequent therapy received	VEGF-targeted therapy (69/263; 26.2%)	Nivo-based or PD-(L)1 inhibitor-based regimen (101/288; 35.1%)
Median TTNT*	[redacted]	[redacted]

Abbreviations: HR, hazard ratio; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; TTD, time to discontinuation; TTR, time to response

Notes: this was not provided in line with the EAG requested definition and was instead defined as (1) the survival time from end of therapy in patients who never received subsequent systematic treatment, and (2) the time from end of therapy until subsequent systematic treatment in patients who received subsequent systematic treatment. An event was defined as receiving subsequent systematic treatment.

Data taken from the company submission and Burotto 2023[120]

By means of comparison, considering earlier follow-up points for the company's primary outcome, PFS rates were: 79.6% versus 59.9% at six months, 67.9% versus 48.3% at nine months, 57.8% versus 37.6% at 12 months, and 37.8% versus 21.7% at 24 months, for cabozantinib + nivolumab and sunitinib, respectively.

Subgroup analysis is provided by the company for a range of factors, including IMDC baseline prognostic risk, which was considered by the EAG to be the most pertinent subgroup analysis. Results were categorised by 0 (favourable), 1-2 (intermediate) and 3-6 (poor) and are presented below in Table 24. Combined intermediate/poor data were also provided for certain outcomes. In particular, it is notable that findings for OS [redacted] in favourable risk patients, in contrast to findings for patients with intermediate and poor risk. While the median OS had not yet been reached in the cabozantinib + nivolumab arm, there was a similar rate in mortality by the final follow-up (cabo+nivo: 30/74 [40.5%]; suni:27/72 [37.5%]). In addition subgroup analysis found [redacted] in the favourable risk group in HRQoL measured by the FKSI-19 with quality of life declining from baseline in both risk groups.[121] This [redacted] creates uncertainties in generalisability and in decision risk.

[redacted]

Table 24: Key 44-month results in CheckMate 9ER by IMDC prognostic risk status

Outcome	Favourable N =74 Int, 72 Con	Intermediate N = 188 Int, 188 Con	Poor N = 61 Int, 68 Con
Median PFS (95% CI)	Int: 21.42 (13.08-24.71) Con: 13.86 (9.56-16.66)	Int: 16.59 (11.86-20.04) Con: 8.67 (7.00-10.38)	Int: 9.92 (5.91-17.56) Con: 4.21 (2.92, 5.62)
Hazard ratio PFS (95% CI)	0.72 (0.49-1.05)	0.63 (0.49, 0.80)	0.37 (0.24, 0.57)
Median OS (95% CI)	Int: N.A. (40.67-N.A.) Con: 47.61 (43.63, N.A.)	Int: 49.48 (37.55, N.A.) Con: 36.17 (25.66, 45.96)	Int: 34.84 (21.36, N.A.) Con: 10.51 (6.83-20.63)
Hazard ratio OS (95% CI)	1.07 (0.63-1.79)	0.75 (0.56-1.00)	0.46 (0.30-0.72)
ORR % (95% CI)	Int: 67.6 (55.7, 78.0) Con: 45.8 (34.0, 58.0)	Int: 56.4 (49.0-63.6) Con: 27.7 (21.4-34.6)	Int: 41.0 (28.6-54.3) Con: 10.3 (4.2, 20.1)

Abbreviations: Int = intervention, cabozantinib with nivolumab. Con = control, sunitinib.

Treatment-related adverse events occurred in 97.2% patients receiving cabozantinib + nivolumab and 93.1% of patients receiving sunitinib with 66.9% versus 55.3% at Grade 3 or higher respectively. Treatment-related AEs led to discontinuation of either nivolumab or cabozantinib in 27.5% of patients versus 10.6% of patients in the sunitinib arm and [redacted] versus [redacted] of patients had at least 1 dose reduction of cabozantinib and sunitinib respectively. The most common treatment-related AEs were diarrhoea, HFS, hypertension, fatigue and hypothyroidism in both arms. Most immune-mediated AEs were low grade and hypothyroidism was the most common immune-mediated AE in both arms; 21.9% of patients treated with cabozantinib + nivolumab required corticosteroids (≥ 40 mg prednisone daily or equivalent) to manage immune-mediated AEs.

Analysis of HRQoL data collected via the FKSI demonstrated a benefit for cabozantinib + nivolumab on the FKSI-19 DRS-v1, 3·48 (1·58–5·39) and EQ-5D-3L UK utility index, 0·04 (0·01–0·07), reaching significance at most timepoints, with small to moderate effect sizes (0·2–0·5).[122] Patients were less likely to be bothered by side effects of for cabozantinib + nivolumab regardless of risk (intermediate / poor-risk odds ratio [OR], 0.50; 95% CI, 0.34–0.75; favourable-risk OR, 0.51; 95% CI, 0.28–0.91).[121] This analysis, however, needs to be considered in the context of the higher rates of discontinuation and dose reduction seen for cabozantinib + nivolumab.

3.5. Description and critique of the evidence presented by other stakeholders

3.5.1. Professional organisation submission

One professional organisation submission was received from the British Association of Urological Surgeons (BAUS).

The submission highlighted that the aim of treatment for RCC varies by disease stage (during Stages 1 to 3c, the aim is to cure, while for Stage 4 disease, the aim is to prolong life of a high quality). BAUS noted that the pathway of care for RCC was not well defined, and there was variation in treatment across different centres. The exact systemic anti-cancer therapy (SACT) and the sequence of treatments used at different points in the pathway will vary from centre to centre as there is currently no predictive tool/marker for each agent. This variation has been established by a recent NHS England-related audit commissioned by Kidney Cancer UK, and it

will soon be illustrated on a yearly basis by the National Kidney Cancer Audit. Nevertheless, the pathway presented in the NICE final scope was considered broadly representative of clinical practice in the UK. The submission highlights European Association of Urology (EAU),[123] European Society of Medical Oncology (ESMO),[43] and American Society of Clinical Oncology (ASCO)[44] guidelines as guidelines used in the treatment of the condition and noted that NICE has commissioned a guideline[124] in this area. Commissioning policies relevant to treatment may exist, but none were specified in the submission.

The BAUS considered that cabozantinib + nivolumab would be a welcome additional 1st line IO/TKI option in addition to the existing avelumab + axitinib combination and considered this to be the most likely treatment replaced as it is thought to be less effective than some of the other combinations. The EAG assumed that this comment related particularly to favourable risk patients for whom avelumab + axitinib was the only option available (via the CDF). They considered cabozantinib + nivolumab to represent only a “marginal gain”. They were not able to provide input to many of the questions related to the specifics of the technology and its impact compared to current care as they did not consider this to be their area of expertise.

3.5.2. Patient organisation submission

One patient organisation submission was received from Action Kidney Cancer.

The submission highlights that living with aRCC/mRCC presents significant challenges for patients and their families. The disease and the side effects of current treatments can have a profound impact, causing financial pressures, emotional distress, and a loss of confidence. Nephrectomy, a common treatment option, carries potential complications and requires a lengthy recovery period. Living with reduced kidney function can lead to long-term complications such as hypertension and chronic kidney disease. It is crucial to provide patients with treatment choices and maintain control to address these burdens effectively.

Family members and caregivers are noted to play a crucial role in supporting patients with aRCC/mRCC but face their own challenges, including financial burdens and the impact of frequent clinic visits. Access to treatments beyond the 1st line is complex and limited, leaving some patients with BSC as their only option. Improved access to new drugs, psychological support services, timely scan results, clinical nurse specialists, and personalised care plans is necessary to enhance overall care and patient experience.

The current treatment pathway for RCC in the UK was described as involving surgery or ablation for early-stage tumours, followed by adjuvant treatment with pembrolizumab where applicable. For aRCC/mRCC, a combination of immunotherapy or targeted therapies is administered. However, these treatments often come with significant side effects that impact the quality of life of patients and their families. Limited access to innovative cancer treatments in the UK may lead to poorer outcomes compared to other regions. Current treatments available on the NHS have their disadvantages, including toxicity, tolerability issues, debilitating side effects, frequent hospital visits for infusion sessions (e.g. with nivolumab), and additional medications for managing side effects. There are limited treatment options beyond the 4th line, which results in patients relying on BSC, with disease progression becoming inevitable.

Certain subtypes of RCC were noted to have poor prognoses and limited treatment responses (e.g. papillary RCC and RCC with sarcomatoid features), highlighting the need for better treatment options. In this regard, the EAG also noted the organisation's reference to hereditary renal cancer and renal medullary carcinoma but noted that these conditions were outside of the scope of this appraisal. Action Kidney Cancer highlight that the UK's cancer survival rates, including kidney cancer, lag behind those of other countries. Access to novel treatments is crucial to improve outcomes and reduce premature deaths. The absence of biomarkers for treatment selection emphasises the need for treatment alternatives in all disease stages. Offering a choice of treatments based on individual patient characteristics and needs is essential for disease management and maintaining quality of life.

The cabozantinib + nivolumab combination was considered to have shown promising results in the treatment of aRCC, improving survival rates and quality of life. Aligned with earlier clinical feedback to the EAG, the cabozantinib + nivolumab combination may be less beneficial for patients with significant co-morbidities or pre-existing autoimmune conditions, such as cardiovascular disease, hypothyroidism, or ulcerative colitis. Access to this combination is restricted for patients with these conditions due to the potential for SAEs or exacerbation of existing health issues. Immunotherapy like nivolumab can worsen autoimmune conditions, requiring lifelong treatment with IV immunosuppressants.

The organisation concludes that addressing the challenges of access, side effects, and limited treatment options is crucial to provide the best possible care and outcomes for patients with aRCC in the UK.

3.6. Critique of real-world evidence identified for this appraisal

3.6.1. Identified real-world evidence

The search and screening process for RWE is described in Section 3.1.2.

A total of four relevant databases were identified in the review of RWE (Table 26). Of these, data were only publicly available for the National Cancer Registration and Analysis Service [NCRAS] [#1][19] database. These data were included. Three databases (Systemic Anti-Cancer Therapy [SACT] dataset [#2][125]; Clinical Practice Research Datalink [CPRD] [#3][126]; Hospital Episode Statistics [HES] [#4][127]), were excluded as data were not available in the public domain and it would not have been possible to acquire the data within the necessary timeframe for this appraisal.

A total of 12 published reports that contained details of potentially relevant data sources were included for additional follow up to request access to data sets (Table 26). The authors for each of the 12 published reports containing potentially relevant data sources were contacted for access to additional data. A three-week period was allowed for a response, with one follow-up email sent. A total of four studies were excluded: four (Marchioni 2021 [#6];[128] International mRCC Database Consortium [IMDC, #7];[129] Schmidinger 2020 [#10];[34] Maroun 2018 [#8][130]) were excluded on geographical location as they reported data for non-UK participants and despite follow-up with the authors UK data could not be obtained, and one study (Olsson-Brown, 2020 [#15][131] was excluded on population as it reported data for a mixed population and data for the 335 participants with RCC could not be obtained from the corresponding author. A total of seven analyses were included: RECCORD (Wagstaff 2016)[22]; UK RWE 2022[53]; Nathan 2022 [132]; Brown 2021[133]; Hack, 2019[134]; Hawkins 2020[32]; NICE TA780[48]).

In addition to the data sets and studies identified in the EAG's review, a further four potential sources were identified in stakeholder and company submissions (Table 25). In addition, to these sources the company also provided hospital audit data 2022 from the same data set reported in Maroun (2018)[130] in its response to clarification question A1 (Table 25). Following scrutiny against the EAG's PICOS criteria specified in Section 3.1.2.3, two studies were excluded on geographical location as they did not report data for UK participants: one study was conducted in Germany (Hilser 2023) and one study was a multicentre study in 32 worldwide institutions (Santoni 2022). Three studies were included that met the specified PICOS criteria (Kidney Cancer UK: Quality Performance Audit of kidney cancer services in England[135]; Nathan

2023;[136] IQVIA hospital audit data 2022[137]). Given that no real-world evidence was identified evaluating the cabozantinib + nivolumab combination, the geographical criterion was relaxed to include the Hilser (2023)[138] study.

In total, 12 sources[19,22,32,48,53,132-138] A summary of the information sources scrutinised is provided in Table 25.

Table 25: Identified potential sources of real-world evidence

#	Name	Identified from	Included
Databases			
#1	National Cancer Registration and Analysis Service (NCRAS)[19]	Web search + Health Data Research UK Innovation Gateway	Yes. Publicly accessible data for the advanced RCC (aRCC) population.
#2	Systemic Anti-Cancer Therapy (SACT) dataset[125]	Web search + Health Data Research UK Innovation Gateway	No. Data that would be required from the SACT dataset for this project are not available in the public domain and cannot be accessed within the timescales of this project.
#3	Clinical Practice Research Datalink (CPRD)[126]	Web search + Health Data Research UK Innovation Gateway	No. Data that would be required from the CPRD for this project are not available in the public domain and cannot be accessed within the timescales of this project.
#4	Hospital Episode Statistics (HES)[127]	Web search + Health Data Research UK Innovation Gateway	No. Data that would be required from the HES dataset for this project are not available in the public domain and cannot be accessed within the timescales of this project.
Publications			
#5	RECCORD (Renal Cell Carcinoma Outcomes Research Dataset) registry (Wagstaff 2016)[22]	Observational studies search	Yes (full text)
#6	REMARCC (Registry for Metastatic RCC)[128]	Observational studies search	No. Study reported data for North American and European centres. The authors were contacted for data from the UK centres, but not data were provided.
#7	IMDC International mRCC Database Consortium[129]	Observational studies search + web search	No. The authors were contacted for data from the UK centres, but no data were provided.
#8	IQVIA real world oncology cross-sectional survey data (Maroun 2018)[130]	Observational studies search	No. Study published in Maroun 2018[130] reported data for European centres. The authors were contacted for data from the UK centres, but no data were provided. However, the company provided hospital audit data 2022 from the same data set in its

#	Name	Identified from	Included
			response to clarification question A1. These data were included (see below)
#9	UK RWE dataset 2022[53]	Observational studies search	Yes (access to data set). The authors were contacted and access to the dataset was granted following contact with authors of Challapalli et al. Patterns of care and outcomes of metastatic renal cell carcinoma (mRCC) patients with bone metastases (BM): A UK multicenter review [Challapalli 2022][139])
#10	Real-world Experience With Suni Treatment in Patients With mRCC: Clinical Outcome According to Risk Score (Schmidinger 2020)[34]	Observational studies search	No. Study reported data for European centres. The authors were contacted for data from the UK centres, but no data were provided.
#11	Ave + axi in advanced renal cell carcinoma (aRCC): 12-month interim results from a real-world observational study in the United Kingdom (Nathan, 2022)[132]	Observational studies search	Yes (conference abstract)
#12	Cabo and axi after VEGF therapy in patients with aRCC: A retrospective cohort study (Brown, 2021)[133]	Observational studies search	Yes (conference abstract)
#13	Real world experience of nivo therapy in metastatic renal cancer patients: A 3 year multi-centre review (Hack, 2019)[134]	Observational studies search	Yes (conference abstract)
#14	Treatment patterns and health outcomes in mRCC patients treated with targeted systemic therapies in the UK (Hawkins 2020)[32]	Observational studies search	Yes (full text)
#15	Real-world outcomes of immune-related adverse events in 2,125 patients managed with immunotherapy: A United Kingdom multicenter series (Olsson-Brown, 2020)[131]	Observational studies search	Yes. Study reported results for a mixed population; 335 participants had RCC. The authors were contacted for access to the RCC data. The authors were chased but no response was received (Feb to last contact, April). No data were provided.
#16	Information from SACT, collected as part of the CDF managed access arrangement, contained in NICE TA780[48]	During grey literature screening/data extraction	Yes (report)

#	Name	Identified from	Included
Stakeholder submissions (company and other stakeholders)			
#17	Kidney Cancer UK: Quality Performance Audit of kidney cancer services in England[135]	Stakeholder submission	Yes (report)
#18	Real-World Data on Cabo in Previously Treated Patients with mRCC: Focus on Sequences and Prognostic Factors (Santoni, 2019).[140]	Company submission	No. Study reported data for 32 worldwide centres, no data from UK centres reported
#19	Cabo + nivo in adult patients with aRCC or mRCC: A retrospective, non-interventional study in a real-world cohort (Hilser, 2023)[138]	Company submission	Yes. Study reported data for German centres only, no UK centres included in the study. Given the lack of evidence on the cabo + nivo combination the geographical setting criterion was relaxed in respect of this intervention
#20	Real-world treatment sequencing and outcomes in patients with aRCC. The CARINA interim analysis (Nathan 2023)[136]	Company response form	Yes (conference abstract + poster)
#21	IQVIA Hospital Audit Data[137]	Company clarification response to question A1	Yes. The company provided hospital audit data 2022 from the same data set as reported in Maroun 2018[130] in its response to clarification question A1. These data were included

Abbreviations: aRCC, advanced renal cell carcinoma; BM, bone metastases; CDF, Cancer Drugs Fund; CPRD, Clinical Practice Research Datalink; HES, hospital episode statistics; mRCC, metastatic renal cell carcinoma; NCRAS, National Cancer Registration and Analysis Service; NICE, National Institute for Health and Care Excellence; RCC, renal cell carcinoma; RECCORD, Renal Cell Carcinoma Outcomes Research Dataset; REMARCC, Registry for mRCC; SACT, Systemic Anti-Cancer Therapy; TA, technology appraisal; UK, United Kingdom; VEGF, vascular endothelial growth factor

Finally, the NICE team attempted to gain and share access to data generated specifically for this project via a healthcare data analytics company. However, no data were provided in time for the appraisal of cabozantinib + nivolumab. Data are expected to become available during the later phase of this project.

3.6.2. Description and critique of real-world evidence

3.6.2.1. Study characteristics

Available evidence comes from retrospective analyses, longitudinal cohort studies, prospective cohorts, registry data analysis, and audits predominantly from centres in the UK. The study periods vary across studies, but they generally cover a range of years data (2009 to 2022) and as such capture a substantial number of patients and treatment data. The study populations

include people with aRCC/mRCC. Sample sizes ranged from smaller cohorts, such as the Nathan 2022[132] study with an advanced population of 36 patients (N=36), to larger patient populations in the UK RWE,[53] which included 1,319 patients. Interventions assessed in the available evidence typically reflect the NICE recommendations during the data collection periods covered by the included evidence.

The Kidney Cancer UK report[135] provided results from a two-year retrospective audit using data extracted from the National Disease Registration Service (NDRS) pre-COVID-19 pandemic. Incident cases of RCC diagnosed between 1 January 2017 and 31 December 2018 were selected from the National Cancer Registration Dataset (NCRD). A total of 18,640 tumours were selected into the cohort, representing 18,421 distinct patients. The audit was conducted to assess the quality of services and to assess whether there was variation in service and treatment in England. There were six quality performance indicators assessed; of these, three provided information in PICO (post-operative 30-day and 12-month all-cause survival in M0 kidney cancer patients who undergo radical nephrectomy or nephron sparing surgery (NSS) and metastatic kidney cancers should receive SACT or active surveillance).[135]

Hospital audit data (IQVIA 2022[137]) were also provided by the company in response to clarification question A1, these data provide information on volume sales for RCC agents in the UK. Limited descriptive information on the data set was available.

The EAG had access to two data sets:

- The NCRAS dataset[19] provides publicly accessible data for the advanced RCC population. The NCRAS forms part of the National Disease Registration Service (NDRS) in NHS Digital. On 1 October 2021, responsibility for the management of the NDRS transferred from Public Health England (PHE) to NHS Digital. The EAG has extracted publicly available data from the NCRAS, specifically the 'Get Data Out' programme. The 'Kidney' dataset contains information on incidence, treatment rates, survival, routes to diagnosis (and other key outcomes) for patients with malignant kidney cancer in England from 2013 to 2019.
- The UK RWE dataset[53] (access kindly provided by the co-investigators: Amarnath Challapalli, Amit Bahl, Gihan Ratnayake, Ricky Frazer and John McGrane) included 1,319 mRCC participants from 15 UK centres, who commenced 1st line systemic therapies between June 2018 and August 2022. Access to the data set was provided following contact with the authors listed on a conference abstract identified in the searches (Challapalli 2022[139]). The EAG has been able to conduct its own analyses using this data set.

Summary study characteristics are provided in Table 26.

3.6.2.2. Baseline characteristics and risk scores

The included evidence all focused on people with aRCC or mRCC. Median age ranged from 59 years to 68 years[19,22,32,48,53,132-138] which broadly mirrored the populations included in the clinical trials (Table 9). Ten analyses reported sex, in these analyses the majority of participants were male.[22,32,48,53,132-136,138]

Of the 12 analyses, the RECCORD data set[22] included only patients with clear cell histology. Six analyses[32,48,53,132,136,138] included a mix of histologies, but clear cell RCC consistently appeared as the most prevalent histological subtype across the studies ranging from 67% in Hilser [2023][138] to 91% in SACT TA780[48] data. Four[22,53,134,138] of the 12 analyses reported the proportion of participants who had undergone prior nephrectomy; this ranged from 50%[22] to 67.9%[134]).

ECOG PS was reported in five analyses[48,132,133,136,138] and the majority of participants were ECOG PS 0 or 1. The proportion of participants with ECOG PS 0 or 1 ranged from 81% to 89% in four studies,[48,132,136,138] one analysis[133] reported only 20% of participants with ECOG PS 0 or 1. Of note, 8% of participants had missing data in the SACT TA780 data set.[48]

Risk score was reported in eight studies.[32,48,53,132,134,136-138] Risk distribution was measured by a combination of IMDC (or Heng criteria),[48,53,132,134,136-138] MSKCC,[32] risk criteria. For convenience, both sets of risk scoring methods are described as producing risk score classes as ‘favourable’, ‘intermediate’ or ‘poor’. The majority of participants across all studies were assessed as intermediate or poor risk categories for each of the scores used (ranging from 59% in Nathan 2022[132] to 100% in the SACT TA780[48] data set) (Table 27). The proportion of participants assessed as intermediate or poor risk broadly matched that in the clinical trial populations (Table 9).

Baseline characteristics are summarised in Table 27.

Table 26: Summary of study characteristics of included RWE

Study name	Study type	Country (number of centres)	Study period	Population	LOT	Interventions	Outcomes evaluated (per PICOS)
UK RWE 2022[53]	Multicentre UK retrospective analysis; patient level data	UK (17)	01/01/2018 to 23/08/2022	Metastatic (N=1,319)	1L; 2L; 3L; 4L 5L	Cabo; suni; pazo; tivo; nivo; evero; axi; ave+axi; lenv+evero; pem+lenv; cabo+nivo; nivo+ipi; nivo	Risk scores (IMDC); treatment patterns; OS; PFS; treatment discontinuation; TTNT; TTP; costs (information on RDI)
Hawkins 2020[32] Full text	Retrospective (longitudinal) cohort	England (2)	01/01/2008 to 31/12/2015	Metastatic (N=652)	1L; 2L; 3L	1L: suni; pazo; evero; Other 2L: suni; axi; evero; Other 3L: axi; evero; Other	Risk scores (MSKCC); treatment patterns; OS; treatment discontinuation
Wagstaff 2016 (RECCORD)[22]	Registry data (RECCORD). Retrospective non interventional study	UK (7: 5 in England; 1 in Wales and 1 in Scotland)	Mar 2009 to Nov 2012	Metastatic (N=514)	1L; 2L; 3L	1L: suni; pazo; evero; sora; tem; IL-2; IFNα; Other 2L: suni; pazo; evero; sora; tem; IL-2; Other 3L: evero; sora; axi; IFNα; Other	Treatment patterns; OS; treatment; discontinuation; TTNT; TTP
Brown 2021[133]	Retrospective cohort	England (NR)	01/01/2011 to 31/01/2020 (Cancer Analysis System)	Advanced (N=1,485)	2L+[a]	Cabo; axi	Treatment patterns; OS
Hack 2019[134]	Retrospective cohort	England (3)	Feb 2016 and Apr 2019	Advanced (N=109)	2L+[b]	Nivo	PFS; OS
Hilser 2023[138] Conference abstract	Retrospective non-interventional cohort	Germany (8)	NR	mRCC (N=67)	1L	Cabo+nivo	Risk scores (Heng); PFS; OS; TTP

Study name	Study type	Country (number of centres)	Study period	Population	LOT	Interventions	Outcomes evaluated (per PICOS)
Nathan 2022[132] Conference abstract	Prospective cohort	UK (4)	After 1 Aug 2019	Advanced (N=36)	1L	Ave+axi	Risk score (IMDC); PFS; OS
Nathan 2023[136] (CARINA: NCT04957160) Conference abstract + poster presentation	Retrospective, non-interventional cohort using CAS	England (6)	NR	Advanced (N=129) (cabo subgroup N=87)	2L	Any + subgroup analysis of 2L cabo	Treatment patterns; treatment discontinuation
NCRAS 2023[19]	UK Registry data (OS for mRCC collected from 2013 to 2019)	UK (England)	2013 to 2019	Advanced and metastatic (N=18,421)	1L+	Various	OS
IQVIA 2022[137]	Hospital pharmacy audit data	UK (England)	NR	RCC treated patients	1L+	[redacted]	Treatment patterns
Kidney Cancer UK (audit of kidney cancer services in England)[135]	Audit data	UK	1 Jan 2017 to Dec 2018	Advanced and metastatic (N=18,421)	1L+	Various	Post operative 30-day and 6-month all cause survival in M0 kidney cancer patients who undergo RN or NSS; variability in access to SACT for people with metastatic kidney cancer
NICE TA780:[48] SACT data report	Part of TA780 committee papers	England	5 April 2019 & 30 November 2020	Advanced (N=814)	2L	Any post 1L treatment with nivo+ipi	Risk scores (IMDC); treatment patterns; OS; TTNT; treatment discontinuation

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; ave, avelumab; axi, axitinib; CAS, Cancer Analysis System; cabo, cabozantinib; evero, everolimus; IFNα, interferon alfa; IL2, interleukin 2; IO, immunotherapy; ipi, ipilimumab; lenv, lenvatinib; LOT, line of treatment; mRCC, metastatic renal cell carcinoma; NCRAS, National Cancer Registration and Analysis Service; NICE, National Institute for Health and Care Excellence; nivo, nivolumab; NR, not reported; OS, overall survival; pazo, pazopanib; PFS, progression free survival; RCC, renal cell carcinoma; RWE, real world evidence; SACT, Systemic Anti-cancer Therapy; sora, sorafenib; suni, sunitinib; TA, technology appraisal; tem, temsirolimus; tivo, tivozanib; TKI, tyrosine kinase inhibitor; UK, United Kingdom

Notes:

[a]Patients initiating 2L+ cabo (prior axi excluded) or axi (prior cabo excluded)

[b]69/109 (63.3%) received nivo as 2L; 30/109 (27.5%) received nivo as 3L; 9.2% (10/109) as 4L+

[c]Checkpoint inhibitor-based combination therapy as 1st line treatment in UK clinical practice

Table 27: Summary of baseline characteristics of included RWE

Study name	Intervention	Line of treatment	N	Age years, median (range)	Male n (%)	ECOG PS n (%)	Histology (% clear cell; % sarcomatoid)	IMDC (fav; int; poor) n (%)	Prior nephrectomy n (%)
UK RWE 2022[53]	Cabo; suni; pazo; tivo; nivo; evero; axi; ave+axi; lenv+evero; lenv+pem; cabo+nivo; nivo+ipi; nivo	1L: 687(52%); 2L: 415 (32%)[b]; 3L: 168 (13%)[b]; 4L 42 (3%); 5L: 7 (0.5%)	1,319	Mean 64.43 (min 21, max 90; SE 0.28)	936 (71%)	NR	Clear cell: 1,092 (82.8%); chromophobe: 11 (<1%); papillary 69 (5.2%); sarcomatoid 7 (); undifferentiated 6 (<1%); other 53 (<1%); missing/NA 81 (<1%)	Fav 294 (22.3%); Int/Poor 1,016 (77.0%); Missing 9 (<1%)	715 (54.2)
Hawkins 2020[32]	Suni (60.7%) (3.2% switched suni→pazo); pazo (37.7%) (5.7% switched suni→pazo); evero 4 (0.6%); Other 6 (0.9%)	1L	652	Mean 64.84 (SD 10.5)	426 (65.3%)	NR	Clear cell: 518 (79.5%); non-clear cell 70 (10.7%); other 22 (3.4%)	MSKCC: fav 73 (11.2%); int 380 (58.3%); poor 174 (26.7%); missing 25 (3.8%)	NR
	Axi (57.1%); evero (41.9%); suni 1 (0.5%); Other 1 (0.5%)	2L	184	Mean 62.97 (SD 10.3)	124 (67.4%)	NR	Clear cell: 141 (76.6%); non-clear cell 28 (15.2%); other 5 (2.7%)	MSKCC: fav 27 (14.7%); int 77 (41.9%); poor 59 (32.1%); missing 21 (11.4%)	NR
	Evero 13 (72.2%); axi 4 (22.2%); Other 1 (5.6%)	3L	18	Mean 65.06 (SD 8.9)	14 (77.8%)	NR	Clear cell: 13 (72.2%); non-clear cell 4 (22.2%); other 1 (5.6%)	MSKCC: fav 2 (11.1%); int 11 (61.1%); poor 2 (11.1%)	NR

Study name	Intervention	Line of treatment	N	Age years, median (range)	Male n (%)	ECOG PS n (%)	Histology (% clear cell; % sarcomatoid)	IMDC (fav; int; poor) n (%)	Prior nephrectomy n (%)
Wagstaff 2016 (RECCORD)[22]	Suni 404 (78.6%); pazo 60 (11.7%); evero 33 (6.4%); sora 6 (1.2%); tem 4 (0.8%); IL-2 3 (0.6%); IFNα 2 (0.4%); Other 2 (0.4%)[a]	1L	514	Mean 61.6 (SD 10.9)	341 (66.3%)	NR	Clear cell: 514 (100%) (clear cell patients only included in the trial)	NR	257 (50.0)
	Suni 12 (14.8%); pazo 8 (9.9%); evero 43 (53.1%); sora 3 (3.7%); tem 1 (1.2%); axi 4 (4.9%); IL-2 2 (2.5%); Other 8 (9.9%)	2L	81[b]	NR	NR	NR	NR	NR	NR
	Evero 8 (50.0%); sora 1 (6.3%); axi 5 (31.3%); IL-2 1 (6.3%); Other 1 (5.9%)	3L	16[b]	NR	NR	NR	NR	NR	NR
NCRAS 2023[19]	NR	NR	NR[c]	NR[c]	NR[c]	NR[c]	NR[c]	NR[c]	NR[c]
IQVIA 2022[137]	[redacted]	1L+	NR	NR	NR	NR	NR	NR	NR
Kidney Cancer UK (audit of kidney cancer services in	NR	1L+	18,421	68 (58, 77)	11,818 (63.4)	NR	NR	NR	NR

Study name	Intervention	Line of treatment	N	Age years, median (range)	Male n (%)	ECOG PS n (%)	Histology (% clear cell; % sarcomatoid)	IMDC (fav; int; poor) n (%)	Prior nephrectomy n (%)
England)[135]									
NICE TA780:[48] SACT data report	Nivo+ipi	1L	814	61 (NR) <40 to 80+ yrs[a]	596 (73%)	0: 285 (35%); 1: 420 (52%); ≥2: 42 (%); Missing 67 (8%)	Clear cell: 740 (91%); Other 74 (9%)	Int 533 (65%); Poor 281 (35%)	NR
Brown 2021[133]	Cabo	122 (27.7%) received ≥3L Tx	440	62.5 (NR)	258 (58.60%)	0-1: 80 (18.2%)	NR	NR	NR
	Axi	359 (34.4%) received ≥3L Tx	1,045	63.0 (NR)	556 (53.2%)	0-1: 213 (20.4%)	NR	NR	NR
Hack 2019[134]	Nivo	2L: 69/109 (63.3%); 3L 30/109 (27.5%); 4L+ 10/109 (9.2%)	109	59 (NR)	79 (72.5%)	NR	NR	Heng scores: fav 19.41%; int 61.2%; poor 18.3%	74 (67.9)
Hilser 2023[138]	Cabo+nivo	1L	67	67.6 (±30)[d]	42 (62.7)	≤1 51 (76.1)	Clear cell: 45 (67.2)	Fav: 15 (22.4); Int: 33 (49.3); Poor 10 (14.9)	38 (56.7)
Nathan 2022[132]	Ave+axi	1L	36	66.2 (39.8-84.1)	(78%)	0-1: 89%	Clear cell: 72%; Other 25%	Fav 39%; int 42%; poor 17%; unknown 3%	NR

Study name	Intervention	Line of treatment	N	Age years, median (range)	Male n (%)	ECOG PS n (%)	Histology (% clear cell; % sarcomatoid)	IMDC (fav; int; poor) n (%)	Prior nephrectomy n (%)
Nathan 2023[136] (CARINA: NCT04957160)	Cabo 80 (74.8%); suni 14 (13.1%); lenv+evero 1 (0.9%); tivo 3 (2.8%); pazo 3 (2.8%); axi 2 (1.9%); pem+axi 2 (1.9%); ave+axi 1 (0.9%); bev 1 (0.9%)[d]	2L	129	Mean 60 (9.9) [n=96][c]	97 (75.2%)	0: 34 (40.0%); 1: 47 (55.3%); ≥2 4 (4.7%) [n=85]	Clear cell: 75 (77.3%); Mixed clear-cell component 6 (6.2%); non-clear-cell 13 (13.4%); Other 3 (3.1%) [n=97]	Fav 12 (14.6%); Int 53 (64.6%); Poor 8 (15.4%) [n=82]	NR
	Cabo	2L	87	Mean 59.1 (9.8) [n=60][c]	64 (73.6%)	0: 22 (41.5%); 1: 30 (56.6%); ≥2 1 (1.9%) [n=53]	Clear cell: 48 (78.7%); Mixed clear-cell component 3 (4.9%); non-clear-cell 7 (11.5%); Other 3 (4.9%) [n=61]	Fav 8 (15.4%); Int 36 (69.2%); Poor 8 (15.4%) [n=52]	NR

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; ave, avelumab; axi, axitinib; bev, bevacizumab; cabo, cabozantinib; evero, everolimus; IFNα, interferon alfa; IL2, interleukin 2; IO, immunotherapy; ipi, ipilimumab; lenv, Lenvatinib; NCRAS, National Cancer Registration and Analysis Service; NICE, National Institute for Health and Care Excellence; nivo, nivolumab; NR, not reported; OS, overall survival; pazo, pazopanib; pem, pembrolizumab; PFS, progression free survival; SACT, Systemic Anti-cancer Therapy; SE, standard error; sora, sorafenib; suni, sunitinib; TA, technology appraisal; tem, temsirolimus; tivo, tivozanib; TKI, tyrosine kinase inhibitor; UK, United Kingdom

Notes:

[a]<40 yrs: 15 (2%); 40 to 49 yrs: 96 (12%); 50 to 59 yrs: 257 (32%); 60 to 69 yrs: 271 (33%); 70 to 79 yrs: 167 (21%); 80+ yrs 8 (1%)

[b]One additional patient was denoted as receiving 2nd line, 3rd line and 4th line treatment but no treatment type was specified

[c]For each year, patient numbers (population/incidence) were reported and stratified according to stage, age band, RCC type). Median/mean age not provided. Gender split, histology, IMDC risk category, prior nephrectomy not provided

[d]Reported in abstract as median (range)

3.6.2.3. Outcomes

The outcomes reported in the included real-world evidence are summarised in Table 28.

Table 28: Outcomes reported in the RWE

Trial name	Risk scores	OS + prognostic variables	PFS + prognostic variables	TTP	TTNT	Discontinuation	Tx patterns (subsequent Tx)	Health costs	HRQoL
UK RWE 2022	IMDC	X	X	X	X	X	X	X[d]	
Hawkins 2020[32]	MSKCC	X				X	X		
Wagstaff 2016 (RECCORD)[22]		X		X	X	X	X		
NICE TA780:[48] SACT data report	IMDC	X			X	X	X		
IQVIA 2022							X		
NCRAS 2023[19]		X[a]							
Kidney Cancer UK (audit of kidney cancer services in England)[135]		X[b]					X		
Brown 2021[133]		X					X		
Hack 2019[134]		X	X	X[c]					
Hilser 2023[138]	Heng	X	X						
Nathan 2022[132]	IMDC	X	X			X			
Nathan 2023[136] (CARINA: NCT04957160)	IMDC					X	X		

Abbreviations: HRQoL, health-related quality of life; MSKCC, Memorial Sloan-Kettering Cancer Center; NSS, nephron sparing surgery; OS, overall survival; PFS, progression free survival; RN, radical nephrectomy; RWE, real world evidence; SACT, Systemic Anti-cancer Therapy; TTNT, time to next treatment; TTP, time to progression; Tx, treatment

Notes:

[a]OS data yearly records (2013-2019) for Stage 1-4 clear cell RCC and RCC NOS patients with confirmed or unconfirmed diagnoses

[b]Reported post operative 30-day all cause survival in M0 kidney cancer patients who undergo radical nephrectomy (RN) or nephron sparing surgery (NSS) and post operative 12 months all cause survival in M0 kidney cancer patients who undergo RN or NSS

[c]Proportion with disease progression only

[d]Data on relative dosing intensity reported, included as dosing used to inform drug costs

3.6.2.4. Critical appraisal real world evidence studies

The DataSAT was completed for UK RWE (2022),[53] Hawkins (2020),[32] RECCORD (Wagstaff 2016[22]) and SACT TA780.[48] Note that the research team had access to the full data set only for UK RWE (2022)[53] and the remaining assessments were completed based on the publicly available information.

For the remaining studies, no assessment was completed as limited information was reported in the public domain to make a full assessment:

- Brown (2021),[133] Hack (2019),[134] Hilser (2023),[138] Nathan (2022),[132] and CARINA (Nathan 2023)[136] were only available as conference abstracts
- Kidney Cancer UK Audit report,[135] and the NCRAS data,[19] limited access to the data set based on information within reports available online.

The DataSAT assessments for the four appraised datasets[22,32,48,53] are summarised below with detail provided in Appendix L.

Data provenance: Data provenance refers to the documented history and origin of data, including its creation, transformation, and movement throughout its lifecycle. Data for three[22,32,53] of the analyses were derived from retrospective chart reviews conducted in various hospital settings in the UK, specifically focusing on patients with RCC. While specific details regarding data preparation, governance, and management are not provided, it can be inferred that the data collection process was clinically led and aligned with the objectives of the respective studies. Limited information is available on the procedures followed in these aspects.

In contrast, the SACT database served as a data source for one[48] of the analyses. This national database in England collects real-time information reported by NHS Trusts through electronic prescribing systems during patient care. The dataset undergoes regular reviews and updates, indicating ongoing efforts for data management and quality assurance. The SACT team, a part of the NCRAS, manages and ensures the quality of the reported data. Compliance with data protection requirements, such as the Data Protection Act 2018 and GDPR 2016, is ensured. Data submission requires completeness checks and adherence to national standards. Over time, data validation has been improving, although certain fields may still have issues related to ascertainment and completeness.

Regarding geographical settings, the data sources were hospital settings (secondary care) within the UK. The UK RWE (2022)[53] data set included patients from 15 UK hospitals who started 1st-line systemic therapy between January 2018 and August 2022. The Hawkins (2020)[32] analysis included patients who initiated 1st-line systemic therapy in two specific hospitals in Cambridge and Manchester between January 2008 and December 2015. The RECCORD data set (Wagstaff 2016)[22] included patients who began 1st-line systemic therapy from seven hospitals across England, Scotland, and Wales, with data collected between March 2009 and October 2012. The SACT database is a national database in England that collects and manages information about systemic anti-cancer therapy treatment. For the included analysis,[48] data from SACT for patients who received nivolumab + ipilimumab during the period of managed access following the NICE Appraisal Committee recommendation in TA581 were analysed.

It is worth noting that the EAG had access to the authors for the UK RWE (2022) dataset,[53] but no additional documents were available beyond those in the public domain for three of the four analyses,[22,32,48] limiting further insights into the data provenance.

Data quality: Across the UK RWE (2022)[53], Hawkins (2020)[32], RECCORD (Wagstaff 2016[22]), and SACT TA780[48] datasets, the included populations were assumed to be accurate, as they relied on information recorded in reliable medical records. Although specific diagnostic codes were not reported, clear inclusion criteria were stated, ensuring the accuracy of participant selection. The SACT TA780[48] dataset was slightly different to the other three datasets in that it selected participants based on Blueteq® applications for nivolumab + ipilimumab for which data were available in the SACT database (matched cohort SACT data to CDF Blueteq® applications for nivolumab plus ipilimumab between 5 April 2019 and 20 November 2020), and it is assumed that patients met the specified criteria for treatment.[48] In all datasets,[22,32,48,53] the majority of items linked to defining the population; e.g. histology type, previous treatments received were reported to have 100% completeness.

In terms of specific variables, the prognostic score assessed using IMDC or MSKCC risk scores typically showed a high level of completeness, albeit a small proportion of missing data reported in two studies.[32,53]

Similarly for treatments received (1st line and subsequent treatments), these data were considered accurate as the information was taken from medical records and linked prescribing

information. In addition, the data were considered complete as there was no indication of missing data in the datasets[22,32,48,53] among the participants who were recorded as having subsequent treatments.

Standard definitions were consistently used for outcomes such as OS, PFS, TTP, and TTNT, providing consistency and accuracy in measurements across the studies. In the SACT TA780 report in particular,[48] the calculation of OS was clearly reported and included vital status verification, tracing, and follow-up. The medical records were assumed to be accurate sources for determining survival time based on treatment start date. For outcomes which may have included some element of clinician judgment e.g. the assessment of progression, the EAG note there may have been some variability between centres and across studies. In most cases, the assessment was based on assessment of multiple markers, such as radiology, symptomatology, clinical investigation, and therapy changes, although primarily radiological assessment was used to determine progression. Medical records were assumed to be accurate sources for determining survival time relative to treatment start date.

It is important to note that for three studies, the completeness and accuracy assessments for study variables were based on the information reported in the publications. Therefore, the overall data quality assessment is based on the information provided in the studies. Overall, the four datasets[22,32,48,53] exhibited reasonable data quality, with a focus on accuracy, completeness, and were based on reliable data sources. The clear definitions and criteria employed in the studies further enhanced the reliability and robustness of the findings.

Data relevance: The four analyses[22,32,48,53] each included a significant number of patients, with sample sizes ranging from 514[22] to 1,319.[53] All four datasets[22,32,48,53] included data from treatment and monitoring in a UK secondary care setting. In three of the four analyses,[32,48,53] the majority of patients had clear cell histology, while one dataset[22] included only patients with clear cell histology. The majority of patients in the datasets were categorised as intermediate or poor risk[22,32,53] according to the IMDC criteria, with one dataset[48] specifically including only patients with intermediate or poor risk RCC. Sufficient data were reported in respect of the analysis populations for the EAG to conclude that the datasets reflected the appropriate population.

The UK RWE (2022)[53] dataset provided valuable insights into the population of RCC patients starting 1st-line systemic therapy in the UK. The data collection spanned from January 2018 to August 2022 and included comprehensive data from 15 UK centres. These data captured the

most recent routine practice in the NHS, reflecting the use of newer treatments recommended by NICE (**1st line:** cabozantinib TA542;[52] tivozanib (TA512);[51] nivolumab + ipilimumab [TA780 via CDF for the majority of the data collection period 2019 to 2022 TA581/TA780];[48] and avelumab + axitinib TA645 [via CDF];[46] **2nd line:** nivolumab TA417;[58] cabozantinib TA463;[55] and lenvatinib + everolimus TA498[56] [refer to Table 29]). The Hawkins (2020)[32] dataset focused on patients with mRCC and obtained data from two specialist centres in England between January 2008 and December 2015. Similarly, the RECCORD study (Wagstaff 2016[22]) analysed data from seven UK centres, providing insights into treatments and outcomes between March 2009 and October 2012. While the data collection periods for these datasets pre-date the recommendations for many current treatment options, comparing them with the more recent UK RWE (2022)[53] dataset can provide insights into the impact of newer treatments on outcomes and the treatment pathway. The SACT TA780[48] dataset specifically focused on patients who received nivolumab + ipilimumab treatment during the managed access period following the NICE appraisal. The dataset included 814 unique patients who applied for CDF funding, and 99% of them had a treatment record in the SACT database. The collection period covered 2019 to 2022 was also sufficient to capture many of the newer treatments recommended by NICE during that period).

Time-to-event outcomes, particularly OS, were assessed in all analyses.[22,32,48,53] In the SACT TA780 dataset[48] median OS had not been reached, but the follow-up period in SACT allowed for the collection of additional information beyond that captured in the trial period. The follow-up durations for each analysis were otherwise deemed sufficient to capture the specified outcomes beyond the trial period and to gather valuable insights into subsequent treatments.

Sample sizes ranged from 514[22] to 1,319[53] participants. The SACT TA780 dataset[48] provided a flow diagram for participants identified to participants included with reasons for not including participants. None of the analyses[22,32,48,53] conducted a sample size calculation as their primary objective was to collect descriptive information rather than test a specific research hypothesis.

Overall, the included datasets[22,32,48,53] provide relevant information from UK practice in terms of treatment patterns and efficacy outcomes (e.g. OS, PFS, TTNT, discontinuation, dosing information). However, in interpreting the information, it is crucial to consider the changes in the treatment landscape over time, given the differences in treatment pathways between the study periods and the present.

3.6.3. Results from real-world evidence

3.6.3.1. Treatment patterns

Feedback received in the both the professional and patient organisation submissions was that the pathway of care for RCC is not well-defined, leading to variation in treatment approaches across different centres. They noted that there is no established predictive tool or marker for each systemic anti-cancer therapy (SACT), resulting in different treatment sequences at different points in the pathway. A recent audit commissioned by Kidney Cancer UK[135] highlighted this variation, suggests that treatment policy is highly variable. The proportion of patients with metastatic kidney cancer who received SACT (with drugs) was widely inconsistent. When stratified by Cancer Alliance, the proportions of metastatic (M1) RCCs that received SACT one month before to any time after diagnosis ranged from 39.7% (95% CI [33.7, 46.1]) to 70.7% (95% CI [59.6, 79.8]). These variations were broadly similar from one month to four years after diagnosis (the cut off was May 2021).

Seven sources reported information on treatment patterns.

Three analyses reported the range of targeted systemic therapies recommended for use in mRCC patients in the UK across lines of therapy (RECCORD [Wagstaff 2016]; Hawkins 2020; UK RWE). The studies were all UK studies and were aligned with the NICE pathways for advanced and/or metastatic RCC, meaning that the received treatments were consistent with NICE-recommended systemic therapies. The broad time period across the three analyses (2008 to 2022) means that the treatments received in the studies vary relative to NICE recommendations at the time the studies were conducted which explains the differences in treatment practices.

The availability of interventions recommended by NICE during the data collection periods for each of the included studies is provided in Table 29. Drugs were considered to be available at the time of publication of final guidance by NICE either with a recommendation for routine commissioning or a recommendation to the Cancer Drugs Fund.

As noted, the interventions received by participants in the earlier data sets[22,32] reflected the treatments available during the study period; i.e. in both data sets the majority of participants received either sunitinib or pazopanib (78.6% and 11.7% and 60.7% and 37.7% in the Hawkins [2020][32] and RECCORD [Wagstaff 2016[22]] data sets, respectively). Subsequent treatments

were broadly similar in the two data sets with the majority of participants receiving everolimus (53.1% and 41.9% in the Hawkins [2020][32] and RECCORD [Wagstaff 2016[22]] data sets, respectively). The main difference being that a larger proportion of participants received axitinib in the later data set (57.1% vs 4.9% in the Hawkins [2020][32] and RECCORD [Wagstaff 2016[22]] data sets, respectively) reflecting the timing of the NICE recommendations. In 3rd line, the majority of participants received everolimus or axitinib (Table 30).

A summary of treatments used from 1st line to 4th line from three RWE sources (data collection period 2008 to 2022) are provided in Table 30. The EAG had access to UK RWE (2022) which includes data aligned with the majority of NICE recommendations. These data indicate that the following treatments are used at 1st line: avelumab + axitinib (13%), nivolumab + ipilimumab (23%), pazopanib (18%), sunitinib (25%), cabozantinib (9%) and tivozanib (8%) aligned with NICE recommendations. The data indicate a small proportion (5%) of patients are treated with interventions not recommended by NICE (e.g. in clinical trials). At 2nd line the data indicate the majority of patients are treated with cabozantinib (39%) or nivolumab (37%) with a smaller proportion of patients receiving lenvatinib + everolimus (5%) or axitinib (3%) and 16% of patients treated with interventions not recommended by NICE (e.g. in clinical trials). When stratified by risk group, the proportions treated were similar apart from a higher proportion of patients receiving nivolumab + ipilimumab in 1st line treatment in the intermediate/poor risk group as would be expected in line with NICE recommendations. Also of note was that, aligned with clinical feedback to the EAG, the proportion of participants receiving avelumab+axitinib was higher in the favourable risk group relative to the intermediate/poor risk group 21.43% vs 10.33%, respectively). A broader range of interventions were used in later lines with cabozantinib the most common treatment at 3rd line (48%) and axitinib the most common treatment at 4th line (43%). A full breakdown of interventions received in the cohort is provided in Appendix L. The EAG conducted an analysis to show the pathway of care from 1st line to 4th line treatment shown in Figure 9 (data are reported in Appendix L).

Table 29: Availability of interventions recommended by NICE during study data collection periods

Intervention		Suni	Pazo	Evero	Axi	Nivo	Cabo	Cabo	Lenv+ evero	Tivo	Nivo+ipi	Ave+axi	Pem+ lenv
NICE appraisal		TA169[49]	TA215[50]	TA219 → TA432[57]	TA333[54]	TA417[58]	TA463[55]	TA542[52]	TA498[56]	TA512[51]	TA780 (CDF review of TA581))[48]	TA645 (CDF)[46]	TA858[38]
Line of treatment recommended		1L (ECOG PS 0 or 1)	1L (no prior cytokine therapy; ECOG PS 0 or 1)	2L (after prior VEGF)	2L (after 1L TKI or a cytokine)	2L	2L (after prior VEGF)	1L (int or poor risk per IMDC criteria)	2L (after 1 prior VEGF and ECOG 0 or 1)	1L	1L (int or poor risk per IMDC criteria)	1L via CDF	1L (int or poor risk per IMDC criteria)
Published guidance date		2009	2011	2011 → 2017	2015	2016	2017	2018	2018	2018	2019 (via CDF); 2022 (CDF review)	2020	2023
Study	**Data collection period**												
RECCORD (Wagstaff 2016)[22]	Mar 2009 to Oct 2012	Y	Y	Y									
Hawkins 2020[32]	1 Jan 2008 to 31 Dec 2015	Y	Y	Y	Y								
UK RWE 2022	1 Jan 2018 to 23 Aug 2022	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	Y (via CDF)	
Brown 2021[133]	1 Jan 2011 to 31 Jan 2020	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	N (publ Sep 2020)	
SACT TA780[48]	4 Apr 2019 to	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	Y(via CDF)	

	30 Nov 2020												
CARINA (Nathan 2023)[136]	15 Jan 2015 to Sept 2022	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	Y(via CDF)	

Abbreviations: 1L, 1st line; 2L 2nd line; ave, avelumab; axi, axitinib; cabo, cabozantinib; CDF, Cancer Drugs Fund; ECOG PS, Eastern Cooperative Oncology Group performance status; evero, everolimus; IMDC, International Metastatic Renal Cell Carcinoma Database Consortium; lenv, Lenvatinib; NICE, National Institute for Heath and Care Excellence; pazo, pazopanib; pem, pembrolizumab; publ, published; RWE, real world evidence; suni, sunitinib; TA, technology appraisal; tivo, tivozanib; TKI, tyrosine kinas inhibitor; UK, United Kingdom; VEGF, vascular endothelial growth factor

Table 30: Treatments used from 1st line to 4th line across three real world evidence studies

	RECCORD (Wagstaff 2016)	Hawkins 2020	UK RWE 2022
	%	n	%
1L			
Ave+axi	0	0	12.7
Cabo	0	0	8.6
Nivo+ipi	0	0	23.4
Pazo	11.7	37.7	17.7
Suni	78.6	60.7	24.7
Tivo	0	0	7.9
Other	9.8[a]	1.5	4.9
2L			
Axi	4.9	57.1	3.0
Cabo	0	0	38.8
Lenv+evero	0	0	4.6
Nivo	0	0	37.3
Evero	53.1	41.9	
TKI (suni, pazo)	24.7		
Other	17.3[a]	1.0	16.3
3L			
Axi	31.3	22.2	11.2
Cabo	0	0	48.1
Lenv+evero	0	0	13.1
Evero	50.0	72.2	4.2
Nivo+ipi	0	0	0.5
Nivo	0	0	19.6

	RECCORD (Wagstaff 2016)	Hawkins 2020	UK RWE 2022
Pazo	0	0	0.5
Suni	0	0	2.3
Tivo	0	0	0.5
Other	18.5	5.6	-
4L			
Axi	0	0	42.6
Belz	0	0	1.85
Cabo	0	0	14.81
Lenv+evero	0	0	9.26
Evero	0	0	20.37
Nivo	0	0	5.56
Other	0	0	3.7
Suni	0	0	1.85
5L			
Axi	0	0	42.86
Belz	0	0	57.14
Total	0	0	100

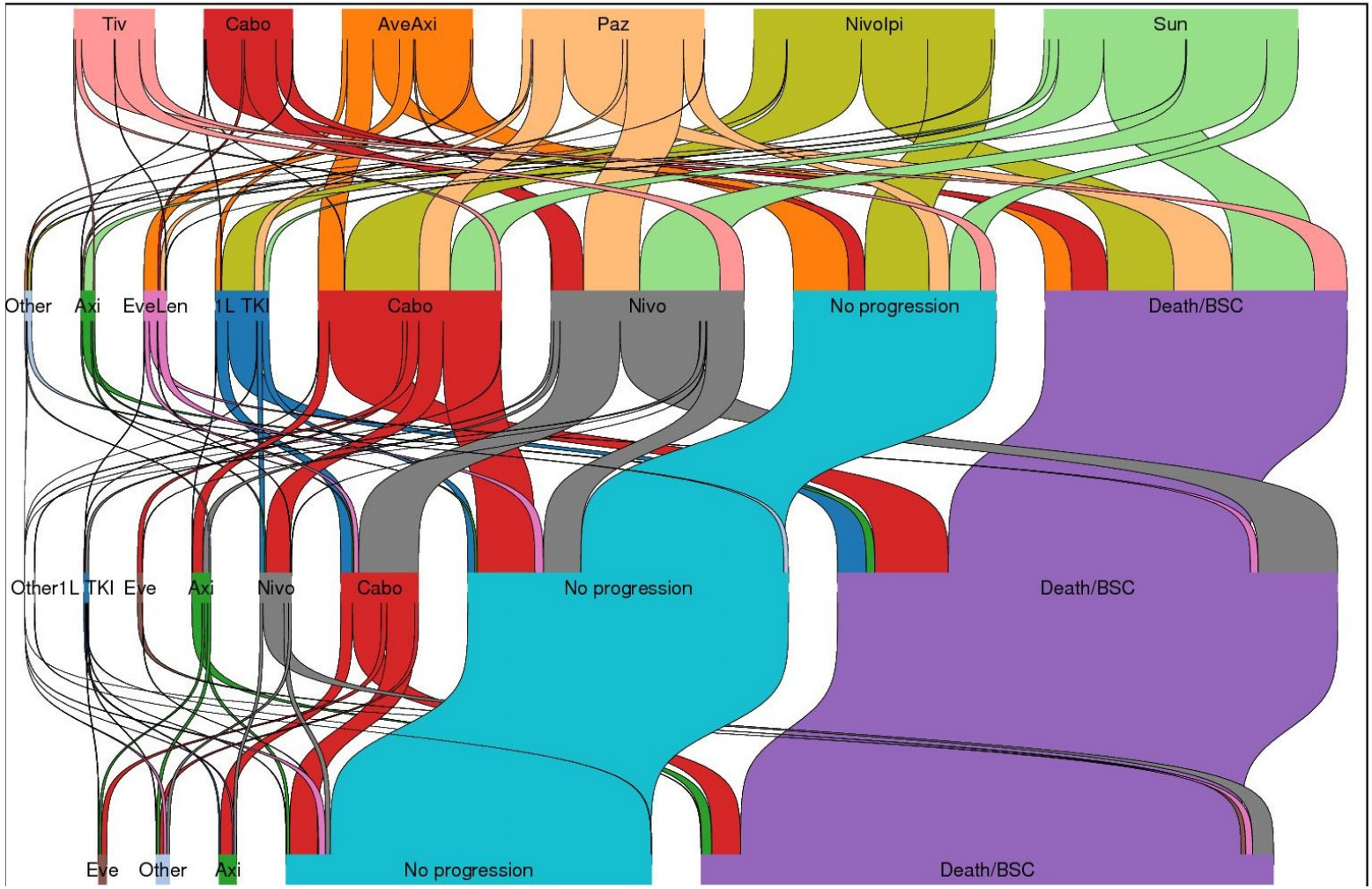
Abbreviations: axi, axitinib; ave, avelumab; belz, belzutifan; cabo, cabozantinib; evero, everolimus; ipi, ipilimumab; lenv, lenvatinib; nivo, nivolumab; pazo, pazopanib; RWE, real world evidence; sora, sorafenib; suni, sunitinib; tivo, tivozanib; UK, United Kingdom

Notes:

aOther grouping included treatments not recommended by NICE or not in the treatment pathway set out in Figure 6: 1L → evero 6.4%; sora 1.2%; tem 0.8%; IL-2 0.6%); IFNα 0.4%; Other 0.4%; 2L → sora 3.7%; tem 1.2%; IL-2 2.5%; other 9.9%

Sources: RECCORD (Wagstaff);[22] Hawkins 2020;[32] UK RWE 2022[53]

Figure 9: Sankey diagram for UK real-world evidence



Abbreviations: 1L, 1st line; Ave, avelumab; Axi, axitinib; BSC, best supportive care; Cabo, cabozantinib; Evero, everolimus; Evero+Len, everolimus + lenvatinib; Nivo+Ipi, nivolumab + ipilimumab; Paz, pazopanib; Sun, sunitinib; TIV, TKI, tyrosine kinase inhibitor

Notes: Patients receiving treatments not currently prescribed in the NHS have been removed from 1st line for readability.

Source: UK RWE 2022[53]

Hospital pharmacy audit data (IQVIA[137]) provided by the company in response to clarification question A1 were provided (Figure 10). These data suggest that [redacted] than other RCC agents in the UK. Although the data do not distinguish between lines of therapy or indication for the different tyrosine kinase inhibitors (TKIs), it can be seen that [redacted]. In addition, the EAG note the [redacted] and a [redacted].

Figure 10: Hospital pharmacy audit data: volume sales by product

[redacted]

Key: Cabometyx = cabo; Fotivda = tivo; Inlyta = axi; Kisplyx = lenv; Sutent = suni; Votrient = pazo
Abbreviations: axi, axitinib; cabo, cabozantinib; lenv, lenvatinib; pazo, pazopanib; suni, sunitinib; tivo, tivozanib
Source: IQVIA 2022[137] (provided by the company in response to clarification question A1

An additional three studies provided information on subsequent therapies following a defined 1st line therapy: two studies described subsequent treatment distributions following 1st line treatment with nivolumab + ipilimumab (SACT TA780[48]; CARINA [Nathan, 2023[136]]); one study described subsequent treatment distributions following treatment with axitinib + avelumab (CARINA [Nathan, 2023[136]]); and, a third study (Brown 2021[133]) described treatment patterns and sequence in patients who received 2nd line-plus cabozantinib or 2nd line-plus axitinib. A summary is provided in Table 31.

Table 31: Sequences described following defined 1st line therapy

	SACTTA780[48]	**Nathan 2023[136] (CARINA: NCT04957160)**		**Brown 2021[c133]**	
N	814	129		440	1,045
1L treatment					
Suni	-	-	-	N=186	N=422

	SACTTA780[48]	**Nathan 2023[136] (CARINA: NCT04957160)**		**Brown 2021[c133]**	
Pazo	-	-	-	N=178	N=500
Nivo+ipi	814 (100%)[a]	107 (82.9%)[b]	-		
Ave+axi	-	-	22 (17.1%)[b]		
Other	-	-	-		
N	**234 (29%)**	**107 (82.9%)**	**22 (17.1%)**	**NR**	**NR**
2L treatment					
Cabo	139 (59.4%)	80 (74.8%)	7 (31.8%)	N=377	0
Suni	31 (13.2%)	14 (13.1%)	1 (4.5%)	0	0
Pazo	28 (12%)	3 (2.8%)	0	0	0
Tivo	19 8.1%)	3 (2.8%)	1 (4.5%)	0	0
Axi	6 (2.6%)	2 (1.9%)	0	0	N=919
Nivo	0	0	2 (9.1%)	0	0
Bev	0	1 (0.9%)	0	0	0
Lenv+evero	5 (2.6%)	1 (0.9%)	10 (45.5%)	0	0
Dabref+tram	2 (0.9%)	0	0	0	0
Pem+carbo	1 (0.4%)	0	0	0	0
Pem+axi	0	2 (1.9%)	0	0	0
Ave+axi	0	1 (0.9%)	0	0	0
Nivo+ipi	0	0	1 (4.5%)	0	0
Evero	1 (0.4%)	0	0	0	0
Irin MDG Panit	1 (0.4%)	0	0	0	0
Trial	1 (0.4%)	0	0	0	0
N				27.7%	34.4%
3L Treatment					
Nivo				N=68	N=171
Axi				N=7	0
Cabo				0	N=49

Abbreviations: ave, avelumab; axi, axitinib; bev, bevacizumab; cabo, cabozantinib; carbo, carboplatin; CDF, Cancer Drugs Fund; dabref, dabrafenib; evero, everolimus; irin, irinotecan; lenv, lenvatinib; MDG, modified de gramont; NICE, National Institute for Heath and Care Excellence; nivo, nivolumab; panit, panitumumab; pazo, pazopanib; pem,, pembrolizumab; SACT, Systemic Anti-Cancer Therapy (data set); suni, sunitinib; TA, technology appraisal; tivo, tivozanib; tram, trametinib

Notes:

[a]Study cohort was participants who had received nivolumab + ipilimumab 1st line in the CDF

[b]Study cohort was participants who had received a 1st line combination therapy including a checkpoint inhibitor

[c]Total for cabo cohort n=440 and total for axitinib cohort n=1,045. The denominator for the reported sequences was unclear from the information available in the conference abstract and data are reported as seen

3.6.3.2. Overall survival

OS was reported in eight sources (Table 32). The studies evaluated various interventions and lines of therapy and typically reported median OS as well as OS rates at different time points, a summary is provided in Table 32.

OS data for RCC were sourced from the NCRAS-published 'Kidney' dataset via the Get Data Out (GDO) platform.[19] These data are reported in Section 2.2.

The Kidney Cancer UK audit report[135] reported post-operative 12-month all-cause survival in M0 kidney cancer patients who undergo radical nephrectomy (RN) or nephron sparing surgery (NSS). A total of 241 (2.8%) of M0 patients who had RN or NSS died in the 365 days after surgery. The most common underlying cause of death for M0 patients who were treated with RN or NSS in the year after their surgery was kidney cancer, accounting for 53.8%. Circulatory disease and other cancers were underlying causes for over 30 deaths each (14.3% and 13.4% of patients respectively).

In the UK RWE (2022[53]) data set, the median OS for patients who received 1st line treatment was [redacted] (95% CI [redacted]) months. The survival estimate was [redacted]% at 12 months falling to [redacted]% at 48 months. For those patients who received a 2nd line treatment, median OS from 2nd line treatment initiation was [redacted] months with a one-year survival estimate of [redacted]%. For those patients who received a 3rd line treatment, median OS from 3rd line treatment initiation was [redacted] months with a one-year survival estimate of [redacted]%. For those patients who received a 4th line treatment, median OS from 4th line treatment initiation was [redacted] months with a one-year survival estimate of [redacted]%. The analysis found that [redacted] (Figure 11). A log-rank test stratifying OS at 1st line by favourable or intermediate/poor status generated [redacted], with a Cox HR of [redacted] (95% CI [redacted]). Refer to Appendix L for Kaplan Meier curves of OS histology, line of treatment, treatment type (by line of treatment), treatment at 1st line by risk category.

Figure 11: UK RWE: Risk stratified overall survival at 1st line

Abbreviations: fav, favourable; int, intermediate; RWE, real world evidence; UK, United Kingdom

Similarly in the Hawkins (2020)[32] analysis median OS decreased with each subsequent treatment. The Hawkins (2020)[32] study found that the Memorial Sloan Kettering Cancer Centre (MSKCC) risk score had a significant impact on OS. Patients with a favourable-risk score had the best survival outcomes, while those with a poor-risk score had the lowest survival outcomes. In both 1st line treatment and 2nd line treatment, significant differences were observed between OS and MSKCC classification ($p<0.001$). At both lines of treatment, favourable-risk patients achieved the best survival outcomes (median OS; 39.7 months [1st line], 14.3 months [2nd line]), compared with intermediate-risk (median OS; 15.8 months [1st line]; 8.9 months [2nd line]), and poor-risk patients (median OS; 6.1 months [1st line] and 3.3 months [2nd line]). The year of treatment initiation also influenced survival, with better outcomes observed for patients treated between 2012 and 2015 (14.2 months) compared to those treated between 2008 and 2011 (11.8 months).

In the RECCORD (Wagstaff 2016[22]) data set, median OS was measured from 1st line treatment initiation and was 23.9 (95% CI 18.6–29.1) months over 13.8 months follow-up. Median OS of patients who received second-line treatment (33.0 months) was significantly longer ($p=0.008$) than that of patients who only received 1st-line treatment (20.9 months. Median OS was significantly longer in participants who switched to 2nd line treatment The authors note that this

may be due to selection bias (good prognosis patients are more likely to receive further therapy), an artefact of the relatively short follow-up period in the study, or because post 1st-line therapy is causing prolongation of survival. pattern was seen when considering the switch to third-line treatment, although it did not reach statistical significance, most likely due to the limited number of patients in this group. In addition, the time interval between diagnosis and systemic treatment was significantly associated with OS ($p<0.001$). Patients who received treatment within 100 days of diagnosis had a lower OS from the start of systemic treatment compared to those who initiated treatment 600 days or more after diagnosis. Toxicity-induced dose decreases also had a significant association with OS ($p=0.002$). Patients who experienced dose decreases in their 1st line treatment had a median survival time of 30.6 months, while for other patients, it was 19.8 months.

The OS observed in the Hawkins (2020)[32] analysis was found to be lower compared to the results reported in the earlier RECCORD database analysis, as well as in the UK RWE (2022)[53] dataset. Several factors could explain the lower median OS observed in Hawkins (2020) when compared to RECCORD and the UK RWE.

Firstly, the RECCORD (Wagstaff 2016)[22]study only included patients with clear cell RCC, which constituted 80% of the cohort in Hawkins (2020)[32] and 82% of the UK RWE dataset. Additionally, the median age of patients in the RECCORD study was younger at 61 years compared to 65 years (mean age) in the UK RWE dataset[53] and 64 years in the Hawkins (2020)[32] dataset. The difference in patient selection and in age distribution could contribute to variations in OS outcomes.

Another potential reason for the lower median OS observed in Hawkins (2020)[32] compared to RECCORD[22] is the inclusion of patients on clinical trials in the RECCORD[22] dataset, as well as a small number of patients receiving IL-2 or IFN-α. Hawkins (2020)[32] suggests that the inclusion of these patients in RECCORD[22] could have contributed to a higher median OS. Hawkins (2020)[32] conducted a subgroup analysis of 89 patients excluded from the main analysis because they received IL-2 or IFN-α at any point during the study. This analysis revealed a substantially longer median OS (47.5 vs. 12.9 months for 1st-line treatment) compared to patients treated exclusively with NICE/CDF-recommended systemic therapies. This discrepancy reflects the fact that the Manchester Centre, where the study took place, is a national treatment centre for high-dose IL-2, which can yield excellent outcomes in carefully selected patients. Furthermore, an additional 72 patients were excluded from the Hawkins (2020)[32] analysis

because they participated in clinical trials where systemic therapies were not administered within the standard of care. These excluded patients could have biased the OS in favour of better outcomes and may partially explain the shorter OS observed in the Hawkins (2020)[32] analysis compared to similar studies.

These differences (patient selection, age, treatment mix) could in part explain the differences between the median OS in the UK RWE (2022)[53] and the Hawkins (2020)[32] dataset, the longer median OS observed in the UK RWE could also potentially be attributed to the availability of newer treatments during the study period. In Hawkins (2020),[32] the majority of participants received sunitinib (60.7%) or pazopanib (37.7%), whereas the UK RWE[53] dataset showed a different distribution with participants receiving avelumab + axitinib (12.7%), nivolumab + ipilimumab (23.4%), cabozantinib (8.6%), tivozanib (7.9%), sunitinib (24.7%), and pazopanib (17.7%) (refer to Section 3.6.3.1 and Table 30).

Overall, the variations in patient selection, age distribution, inclusion of patients on clinical trials, use of specific treatments, and exclusion of certain subgroups can all contribute to the differences observed in median OS between the studies mentioned.

Four other studies reported median OS associated with specific interventions in the aRCC population:

- Nivolumab + ipilimumab as a 1st line treatment showed survival rates at six, 12 and 18 month timepoints of 80%, 69%, and 61%, respectively and median OS was not reached. Sensitivity analysis by IMDC score showed a similar pattern in survival rates at six, 12 and 18 month timepoints gave a median OS of 15 months for IMDC score 3-6 and median OS was not reached in patients with an IMDC score of 1-2;[48]
- Cabozantinib and axitinib as 2nd line treatments demonstrated similar median OS[133] Median OS was lower in RWE than in clinical trials for both cabozantinib (versus everolimus) and for axitinib (versus sorafenib) (Table 13)
- Nivolumab in 2nd and subsequent lines of treatment showed a 12-month survival rate of 56.88%. OS data not reported for CheckMate 025 (median OS not reached) with which to compare (Table 13);[134] and,

Avelumab + axitinib 1st line treatment showed a 12-month OS rate of 86%.[132] OS data not reported for JAVELIN Renal 101 (not estimable) with which to compare (Table 13).

Table 32: Overall survival estimates from RWE

Study	LOT	Intervention	OS definition	N	Median follow-up (95% CI)	Median OS months (95% CI)	OS rate at:
UK RWE 2022[53]	1L	Ave+axi; cabo; nivo+ipi; pazo; suni; tivo	Time from start of 1L treatment to death	1,319	16.8 months (15.8, 17.6)	[redacted]	12 mths: [redacted] 24 mths: [redacted] 36 mths: [redacted] 48 mths: [redacted]
	2L	Axi; cabo; lenv+evero; nivo; pazo; suni; tivo	Time from start of 2L treatment to death	632		[redacted]	12 mths: [redacted] 24 mths: [redacted] 36 mths: [redacted]
	3L	Axi; cabo; lenv+evero; nivo; suni	Time from start of 3L treatment to death	214		[redacted]	12 mths: [redacted] 24 mths: [redacted] 36 mths: [redacted]
	4L	Axi; evero	Time from start of 4L treatment to death	54		[redacted]	12 mths: [redacted] 24 mths: [redacted]
Hawkins 2020[32]	1L	Suni; pazo; evero; Other	Time from the start of 1L treatment to death	652	Mean 23.8 (22.2, 25.4)	12.9 (NR)	12 mths; 52.4% (48.6, 56.4%) 24 mths: 30.9% (27.3, 34.9%) 36 mths: 22.6% (19.3, 26.6%) 60 mths: 10.8% (8.0, 14.6%)
	2L	Suni; axi; evero; Other	Time from the start of 2L treatment to death	184	Mean 21.5 (NR)	6.51 (NR)	12 mths: 31.5% (25.2, 39.5%) 24 mths: 17.0% (11.8, 24.7%) 36 mths: 7.1% (3.1, 16.5%) 60 mths: 7.1% (3.1, 16.5%)
	2L	Suni; axi; evero; Other	Time from the start of 1L treatment to death	184	Mean 21.5 (NR)	20.8 (NR)	NR

Study	LOT	Intervention	OS definition	N	Median follow-up (95% CI)	Median OS months (95% CI)	OS rate at:
	3L	Axi; evero; other	Time from the start of 3L treatment to death	18	Mean 26.1 (NR)	5.91 (NR)	12 mths: 23.8% (10.1, 55.9%); 24 mths: 7.9% (1.3, 48.7%)
	3L	Axi; evero; other	Time from the start of 1L treatment to death	18	Mean 26.1 (NR)	36.7 (NR)	NR
Wagstaff 2016 (RECCORD)[22]	1L; 2L; 3L	As listed for 1L, 2L, and 3L	Time from the start of 1L treatment to death	431	13.1 (12.0, 14.1)	23.9 (18.6, 29.1)	NR
NICE TA780:[48] SACT data report	1L	Nivo+ipi	Time from the start of their treatment to death or censored date	814	3 (NR) (91 days)	Not reached	6 mths: 80% (77, 83%) 12 mths: 69% (65, 72%) 18 mths: 61% (57, 64%)
		Nivo+ipi (≥6 mths follow-up[b])		757	11.9 (NR)	Not reached	NR
		Nivo+ipi (IMDC Int, score 1 or 2)		533	8.7 (NR)	Not reached	6 mths: 88% (84%, 90%) 12 mths: 76% (72%, 80%) 18 mths: 69% (64%, 73%)
		Nivo+ipi (IMDC poor, score 3 or 4)		281	NR	15 (NR)	6 mths: 67% (61%, 72%) 12 mths: 55% (49%, 61%) 18 mths: 45% (38%, 51%)
Brown 2021[133]	≥2L	Cabo	NR	816	NR	11.24 (5.65, 27.98)[a]	NR
		Axi		1,483		10.39 (4.70, 22.03)[a]	NR
Hack 2019[134]	≥1L	Nivo	Time from the start of treatment to death	109	NR	NR	12 mths: 56.88% (NR)
Hilser 2023[138]	1L	Cabo+nivo	NR	67	8.3 (NR)	Not reached	NR
Nathan 2022[132]	1L	Ave+axi	NR	36	12 (NR)	NR	12 mths: 86% (74.8, 97.4%)

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; ave, avelumab; axi, axitinib; cabo, cabozantinib; CI, confidence interval; evero, everolimus; IFNα, interferon alfa; IL-2, interleukin 2; ipi, ipilimumab; mths, months; NICE, National Institute for Health and Care Excellence; nivo, nivolumab; NR, not reported; OS, overall survival; pazo, pazopanib; RWE, real-world evidence; SACT, systemic anti cancer therapy; sora, sorafenib; suni, sunitinib; TA, technology appraisal; tem, temsirolimus; Tx, treatment; UK, United Kingdom

Notes:

Kidney Cancer UK audit report and the NCRAS data reported in a separate table as OS reported by disease stage or post-operative survival rather than by intervention

[a]Propensity score matching (IPW) was used to reduce baseline differences between the cohorts

[b]Sensitivity analyses was also carried out for OS on a cohort with at least six months follow-up in SACT. To identify the cohort, CDF applications were limited from 5 April 2019 to 28 October 2020.

3.6.3.3. Progression-free survival

Four sources reported data on PFS. A summary is provided in Table 33.

The UK RWE (2022[53]) cohort reported a median PFS for 1st line treatment of [redacted] months (95% CI [redacted]) reducing to [redacted] months (95% CI [redacted]) in the cohort of patients receiving 4th line treatment (Table 33). Survival curves for PFS at 1st, 2nd, 3rd, and 4th line are provided in (Figure 12, Figure 13, Figure 14, and Figure 15, respectively). Refer to Appendix L for Kaplan Meier curves of PFS stratified by risk group, histology, treatment, and line of treatment, and post progression survival.

In a retrospective cohort study (Feb 2016 to Apr 2019; England) evaluating nivolumab in 2nd and subsequent lines of treatment (Hack 2019),[134] 31.5% showed a response to nivolumab, 9.3% had stable disease and 59.3% had disease progression. Reported median PFS from the start of nivolumab treatment was 5.4 months (Table 33).

In a retrospective cohort study (Hilser 2023)[138] evaluating patients with mRCC receiving cabozantinib + nivolumab 1st line the PFS rate at six months was 81.9% (Table 33). This was broadly aligned with the rate reported in the CheckMate 9ER trial for cabozantinib + nivolumab (79.6%) (Section 3.3.3.2).

A prospective cohort study (Aug 2019 to Jan 2022; UK) evaluating patients with aRCC receiving avelumab + axitinib 1st line via an early access scheme (Nathan 2022),[132] reported median duration of follow-up and PFS of 12 months (Table 33).

Three sources reported TTP:

- In the UK RWE (2022[53]) data set, median TTP at 1st line was [redacted] months (95% CI [redacted]). The correlation of TTD and PFS (1st line) and TTP (1st line) is [redacted] (Spearman's correlation). Refer to Appendix L for Kaplan Meier curves of time to progression by line of treatment, and for time to progression on 1st line treatment risk stratified.
- In the RECCORD study (Wagstaff 2016),[22] at the time of analysis, disease progression had been experienced by the majority (66.1%) of patients on 1st line therapy (median duration of follow-up: 13.1 months, 95% CI 12.0–14.1 months). Median time to disease progression was 8.8 months (95% CI 7.7–9.9 months). There was a significant association between the time from RCC diagnosis to 1st line treatment and disease progression (p=0.019). Estimated time to progression was shortest for patients who had started 1st line treatment within 100 days of diagnosis (16.8 months [95% CI 14.1–19.5 months]).
- Hack (2019)[134] reported 59.3% had disease progression in the cohort of mRCC patients who received nivolumab in 2nd line-plus treatment. TTP was not reported.

Figure 12: UK RWE: Pooled PFS at 1st line

Abbreviations: RWE, real world evidence; UK, United Kingdom

Figure 13: UK RWE: Pooled PFS at 2nd line

Abbreviations: RWE, real world evidence; UK, United Kingdom

Figure 14: UK RWE: Pooled PFS at 3rd line

Abbreviations: RWE, real world evidence; UK, United Kingdom

Figure 15: UK RWE: Pooled PFS at 4th line

Abbreviations: RWE, real world evidence; UK, United Kingdom

Table 33: Progression-free survival estimates from RWE

Study	LOT	Intervention	Median follow-up	Time on treatment	N	Median PFS mths (95% CI)	PFS rate %
UK RWE 2022[53]	1L	Suni; cabo; nivo+ipi; pazo; tivo	16.8 months (15.8, 17.6)	[redacted]	[redacted]	[redacted]	[redacted]
	2L	Axi; cabo; lenv+evero; nivo; pazo; suni; tivo		[redacted]	[redacted]	[redacted]	[redacted]
	3L	Axi; cabo; lenv+evero; nivo; suni		[redacted]	[redacted]	[redacted]	[redacted]
	4L	Axi; evero		[redacted]	[redacted]	[redacted]	[redacted]
Hack 2019[134]	2L; 3L; 4L+	Nivo	NR	NR	109	5.4 (NR)	NR
Hilser 2023[138]	1L	Cabo+nivo	8.3 (NR)	NR		NR	6 mths 81.9%
Nathan 2022[132]	1L	Ave+axi	12 (NR)	NR	36	12 (NR)	NR

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; ave, avelumab; axi, axitinib; cabo, cabozantinib; CI, confidence interval; evero, everolimus; IO, immune-oncology; ipi, ipilimumab; lenv, lenvatinib; LOT, line of treatment; mths, months; nivo, nivolumab; NR, not reported; pazo, pazopanib; PFS, progression free survival; RWE, real world evidence; suni, sunitinib; tivo, tivozanib; UK, United Kingdom

3.6.3.4. Time to next treatment

Three sources reported time to next treatment (Table 34).

Table 34: Time to next treatment estimates from RWE

Study, year	N	LOT → LOT	Median time (months) to next treatment (95% CI)
UK RWE 2022[53]	1,319 1L → 604 2L	1L → 2L	[redacted]
RECCORD Wagstaff 2016[22]	514 1L → 81 2L	1L → 2L	2009 to 2010: mean 17.4 (SD 11.8) 2010 to 2011: mean 12.3 (SD 7.1) 2011 to 2012 cohort: mean 6.3 (SD 3.7)
SACT TA780[48]	814 1L → 234 2L	1L → 2L	41 days (from last nivo+ipi cycle to next Tx); 148 days (from first nivo+ipi cycle to next Tx)

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; AE, adverse event; CI, confidence interval; evero, everolimus; IFNα, interferon alpha; IL2, interleukin 2; ipi, ipilimumab; LOT, line of treatment; nivo, nivolumab; NR, not reported; pazo, pazopanib; RWE, real world evidence; SACT, Systemic Anti- Cancer Therapy; SD, standard deviation; sora, sorafenib; suni, sunitinib; tem, temsirolimus; Tx, treatment

Notes:

[a]As a percentage of patients who already experienced one dose decrease

[b]Includes n=35 patients who changed to a different 1st line treatment due to toxicity

3.6.3.5. Discontinuation

Five sources reported data on discontinuation (Table 35).

Treatment duration by treatment type at 1st line, 2nd line, 3rd line, and 4th line for the UK RWE data set is provided in Appendix L

Table 35: Discontinuation estimates from RWE

Study, year	LOT	N	Median follow-up mths (95% CI)	Discontinuations, n (%)	Median TTD (months) to discontinuation (95% CI)	Reason for discontinuation n (%)
UK RWE 2022	1L	1,319	16.8 months (15.8, 17.6)	[redacted]	Treatment duration by treatment type at 1L in Appendix L	[redacted]
	2L	604		[redacted]	Treatment duration by treatment type at 2L in Appendix L	[redacted]
	3L	202		[redacted]	Treatment duration by treatment type at 3L in Appendix L	[redacted]
	4L	48		[redacted]	Treatment duration by treatment type at 4L in Appendix L	[redacted]
Hawkins 2020[32]	1L	652	23.8 (22.2, 25.4)	574 (88.0)	10.5 (9.5, 11.6)	Disease progression 411 (71.6); treatment toxicity/ AE 108 (18.8); Other 106 (18.5)
	2L	184		159 (86.4)	5.2 (4.2, 6.3)	Disease progression 115 (72.3); treatment toxicity/ AE 31 (19.5); Other 33 (20.8)
	3L	18		16 (88.9)	5.6 (1.7, 9.5)	Disease progression 11 (68.8); treatment toxicity/ AE 5 (31.3); Other 2 (12.5)
Wagstaff 2016[22]	1L	514	13.1 (12.0, 14.1)	97 (18.9) [b]; 27 (17.1)[a]	4.0 (0.2–5.8) (time to discontinuation of a 1st line drug)	NR
	2L	81		12 (14.8); 0 (0)	NR	NR
	3L	16		2 (12.5); 0 (0)	NR	NR

Study, year	LOT	N	Median follow-up mths (95% CI)	Discontinuations, n (%)	Median TTD (months) to discontinuation (95% CI)	Reason for discontinuation n (%)
SACT TA780[48]	1L	814	3 (NR)	NR	NR	At end of treatment: 469 (58%) stopped treatment: Died not on treatment 131 (28%); disease progression 128 (27%); toxicity 94 (20%); no treatment in at least 3 mths 65 (14%); died on treatment 24 (5%); completed as prescribed 23 (5); patient choice 2 (<1%); COVID 2 (<1%)
Nathan 2022[132]	1L	36	NR	5	NR	Disease progression 4 (11); toxicity 1 (3)
CARINA Nathan 2023[138]	1L	118	NR	NR	10.2 weeks (9.1, 17.1)	NR
	1L subgroup of cabo 2L	83	NR	NR	9.1 weeks (8.1, 12.0)	NR
	2L	129	NR	NR	23.6 weeks (14.0, 28.3)	NR
	2L cabo subgroup	87	NR	NR	28.1 weeks (20.1, 37.1)	NR

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; AE, adverse event; CI, confidence interval; LOT, line of treatment; mths, months; NR, not reported; RWE, real world evidence; SACT; Systemic Anti-Cancer Therapy (data set); TA, technology appraisal; TTD, to discontinuation

Notes:

[a]As a percentage of patients who already experienced one dose decrease
[b]Includes n=35 patients who changed to a different 1st line treatment due to toxicity

3.6.3.6. Health-related quality of life

None of the included real-world evidence studies reported HRQoL.

3.6.3.7. Costs

None of the included real-world evidence studies reported costs.

The UK RWE did report data that enabled the calculation of relative dosing intensity (RDI) which can be used to calculated drug costs, these data are provided in Appendix L.

3.7. Indirect comparisons

3.7.1. Methods

RCTs were synthesised using appropriate meta-analysis methods. Evidence networks for each outcome were formed by decision point on the pathway (i.e. line of treatment or class of prior treatment), combining 2nd, 3rd and 4th line RCC due to trials generally including patients who were previously treated at multiple lines and similar comparator sets.

The feasibility of network meta-analyses (NMAs) was considered by examining where possible the distribution of likely effect modifiers over the networks. Clinical advisors highlighted IMDC prognostic risk category, histology (though information is limited to clear cell vs non clear cell), whether the patient had a prior nephrectomy, and sarcomatoid features (discussed in Section 2.3). We further considered trial results (including interactions in forest plots), any relevant discussion from TA858, and information in the company submission. Due to clinical salience and consistency (and inconsistency) of reporting, we focused on risk, age, line, bone metastases, sarcomatoid features, prior nephrectomy and histology as key effect modifiers, including line where trials included combinations of treated and untreated patients. We did not judge that the feasibility of any NMAs was precluded, but note that relatively sparse evidence networks precludes formal testing via e.g. meta-regression for differences between groups, and consider how analyses might have been impacted by distribution of effect modifiers across the network (see Section 3.7.2.2). In some proportional hazards NMAs in 1st line, we sensitivity analysed findings excluding trials that did not enrol poor-risk patients, partly because several trials suggested that TKIs were not differently effective from more modern (IO or IO combinations) in favourable-risk patients.

Separate networks were formed by line of treatment (1st line or 2nd line-plus) and for 1st line treatment further stratified by IMDC risk subgroup.

If the network contained a clear reference treatment (placebo or standard of care or a central node) then baseline risk was compared across trials using PFS in the reference treatment. The baseline risk serves as a rough proxy for treatment effect modifiers across the trials, some of which may not have been measured or collated. Heterogeneity in baseline risk may point to variation in the distribution effect modifiers over the network, and therefore potential bias in network-based treatment effect estimates.

The set of selected trials from the search process (Section 3.3.1 and 3.3.2) were processed according to Steps two and three of the algorithm outlined by Dias et al.[141] namely: (2) identify all the trials that compare two or more comparators in the population of interest (3) remove trial arms that are not comparators of interest from trials with more than two arms.

Where necessary, connecting nodes were introduced which function to connect networks but do not in themselves represent comparators of interest, similar to the process in TA858.[38] As described above, these nodes were sorafenib and placebo.

NMAs were carried out for the following time-to-event outcomes: PFS and OS. Investigations on the feasibility of time-to-event NMAs for time-on-treatment and time-to-next-treatment indicated insufficient studies available.

Continuous and binary outcomes were further grouped with respect to similarity of follow-up times and combined using odds ratios, as appropriate. Time to event outcomes were analysed using two strategies: one primary and one exploratory. The exploratory strategy, for all time-to-event outcomes, relied on hazard ratios from longest follow-up combined after log transformation using an inverse variance method. We also describe these as ‘proportional hazards NMAs’.

The primary strategy, which focused on PFS as a priority outcome, used a parametric modelling method. OS was included as a secondary outcome. PFS was defined as the time from treatment initiation to the first of RECIST-defined progression or death assessed by BICR, with IA-assessed PFS used if BICR was not available.

3.7.1.1. Fractional polynomial NMAs

The first strategy used fractional polynomial analyses as, based on previous appraisals in RCC, it is expected that there may be issues in justifying proportional hazards for all endpoints. Model selection compared second-order fractional polynomials (except 'repeated powers') drawn from the set of powers defined by −2, −1, −0.5, 0, 0.5, 1, 2, 3 as standard.[142]

Pseudo-individual patient data (IPD) data for survival were requested from the submitting company who provided PFS and OS data for a subset of the EAG network. Further curves were digitised by the EAG. Grouped survival data were then formed in time intervals. The EAG attempted to use the planned grouping interval for survival data of one week (consistent with the model cycle length) but model fits were poor. The EAG elected to use eight weeks in order to obtain stable results and reduce coding manipulations (two months is the value coded by Wiksten[143]).

Initial fractional polynomial model selection used frequentist fixed effects models, identifying a candidate set of 'most likely' models on the basis of visual fit to observed data, clinical plausibility including elicited landmark survival estimates and biological considerations and statistical fit using Akaike Information Criterion (AIC).[143] Frequentist code was largely based on that provided by Wiksten.[143] The selected fractional polynomial model(s) were submitted to Bayesian analysis in the next stage.

A Bayesian analysis of selected models was carried out introducing random effects and comparing these to fixed effects models. Random effects were only be considered on the basis of 'time-invariant' heterogeneity, that is only using between-study variance on intercept terms.[142] The general framework used random effects in a Bayesian framework with Markov chain Monte Carlo estimation, including informative priors from Turner (2015)[144] where available and appropriate and vague or weakly informative priors otherwise. Turner 2015 offers priors for a set of generic scenarios in healthcare and associated types of outcomes. Specifically, an informative prior for the variance of LN(-3.95, 1.79^2) was used, which Turner offers for pharmacological vs pharmacological comparisons with outcomes relating to cause-specific mortality, major morbidity event and composite (mortality or morbidity) outcomes.

Estimation used two chains of 100,000 iterations with 20,000 iterations discarded as burn-in and thinning to every 10^{th} value. Bayesian model comparisons used Deviance Information Criterion

(DIC). Convergence was assessed using standard methods, including autocorrelation and Brooks-Gelman-Rubin diagnostic plots.

Bayesian coding utilised the gemtcPlus package.[145] Fitted curves were compared to the life-table estimates of the hazard following the equation given by Collett p29.[146]

To summarise, each fractional polynomial analysis fits 28 models under any risk and prior treatment subgroup, see for example Table 39 for the case line 1 PFS all risk. Any model selected from these fits is further fitted with fixed effect or random effect alternatives in a Bayesian analysis. An informed selection from these numerous models was made combining statistical criteria (selecting on the basis of smaller AIC or DIC) with clinical or logical plausibility. The steps were:

- Calculate AIC for all FP models with frequentist, fixed effects (FE) approach
- Select models with delta AIC≤5
- For each selected model, run Bayesian models (FE and random effects [RE]) and calculate
  - DIC
  - area under survival curve up to horizon (i.e. restricted mean survival time, or RMST)
- Select models where RMST > threshold for every treatment curve
- Select models best conforming to expert elicitation landmark distributions
- Select model with minimum DIC comparing random and fixed effects

Under expert elicitation the expected survival at five years (conditional on surviving to three) and 10 years (conditional on surviving to 5) were calculated for each model curve for the 1L intermediate/poor risk and 2nd line+ populations. These were compared with the elicitation distributions (Section 4.2). A good match to the expert elicitation was considered to be obtained when the point estimate for the FP NMA conditional survival fell within the 95% confidence interval of the expert elicitation result for that treatment. Models were selected where possible to maximise concordance with the expert elicitation noting that this was not possible in some cases.

Calculation of survival curves involved integration of the modelled hazard using the gemtcPlus package. Unstable results were obtained when the lower integration limit was set to near zero. The EAG attributes this to ‘end effects’ of fractional polynomials including singularities at 0 when exponents are negative. The EAG understands that the relevant gemtcPlus function effectively

applies a constant and finite initial hazard over a width determined by the user. The EAG set this to two weeks to avoid implausibly low survival curve estimates.

3.7.1.2. Proportional hazards NMAs and NMAs of other outcomes

Finally, meta-analyses on proportional hazards estimates were undertaken of survival outcomes, overall response rate, discontinuation due to adverse events and the risk of treatment-emergent adverse events of grade 3 or higher. The EAG also undertook a sensitivity analysis conducted using IA where available for the latest datacut. For trials which compared sequences of treatments, only the first treatment within the sequence was included within the analysis. Thus, for OS, the three relevant crossover trials (SWITCH, SWITCH II and CROSS-J-RCC) were excluded from the 1st line NMA. This is because (i) the results appeared to be reported as HRs for the difference between treatment sequences rather than between treatments (ii) as mentioned the crossover trials served only to connect tivozanib to the main network, and previous technology appraisals considered that an assumption of similar effectiveness to sunitinib was appropriate.[38,46]

The EAG used a Bayesian framework with 100,000 iterations per chain after 10,000 burn-in iterations were discarded and the resultant estimates thinned by using every tenth iteration. We used standard inconsistency and convergence checks on these models.

3.7.2. Characteristics and appraisal of trials identified and included in the indirect comparisons

The majority of included trials were associated with either 1st line or 2nd line-plus populations, but in one prioritised trial, TIVO-1,[103] the study population was mixed. In both cases analyses by line of treatment were available.

Networks were formed for 1st and 2nd line-plus treatments for the outcomes OS, PFS and ORR, taking into account availability of information (as HR, KM curves or response rates), and at 1st line for two IMDC risk categories: intermediate/poor and favourable. Network diagrams for 1st line PFS and OS (all risk) are shown in Figure 16 and Figure 17. Other networks in draft form are supplied in Appendix E.

Many networks are not complete. Following the precedent in TA858 and other previous RCC appraisals, two treatments (sorafenib, placebo) were introduced as connecting nodes. At 1st line, for PFS (Figure 4), this connects tivozanib and results in a complete network, but for OS

(Figure 5), tivozanib is excluded. This is in line with TA858 where the EAG considered that it was not possible to connect tivozanib to the OS network as the OS data required to connect the TIVO-1 trial came from crossover trials (CROSS-J-RCC, SWITCH and SWITCH II) which were not considered suitable as patients switched to the treatment they did not initially receive on progression. This is not considered to be a major issue given that the base case model structure does not use 1st line OS data and previous appraisals have considered that tivozanib is at best similar to pazopanib and sunitinib (TA858, TA645). The full results for excluded treatments with and without these connecting nodes are shown in the table in Appendix E.

For line 2+ networks under FP analyses, the BERAT trial was removed from the network; and indeed BERAT was only helpful for some network meta-analyses in other outcomes. The BERAT trial gives uninformative estimates of treatment effect (PFS HR for everolimus vs TKI was 1.0 (0.26 to 3.85) and OS HR was 1.12 (0.27 to 4.61)) relating to the small trial size (n=10). Inclusion of the trial caused instability in the FP NMA results. This trial also contains some design/reporting flaws, including lack of clarity about design (crossover or parallel group), no protocol available, no power calculation, and an apparent ad hoc extension beyond the planned treatment of axitinib to the class of TKI inhibitors (see the Clinical Trials Registry record for more details[147]). There are two corollaries: that (i) Inference to treatment with axitinib is lost, and that (ii) TIVO-1, TIVO-3 and AXIS trials are also removed, though these latter are not associated with treatments of primary interest. Similarly, for NMAs using proportional hazards and for other outcomes, our analyses relied substantially on the inclusion of BERAT as a linking trial between two components of the network: one defined by everolimus, nivolumab, placebo, everolimus with lenvatinib, and cabozantinib; and another defined by axitinib, sorafenib and tivozanib. This was an imperfect solution given the small size of the trial (n=5 in each arm) and documented issues with protocol administration. For ORR and discontinuation, problems with the data in BERAT (i.e. lack of events in one or both arms) meant that we could not connect both network components. In these analyses, we only present results for the first network component. We also had a disconnected network in our analysis for grade 3 or higher treatment-emergent adverse events, described below. Within subsequent cost-effectiveness analysis given the difficulties making comparison to axitinib within the NMA we test the assumption of equivalence with everolimus consistent with previous technology appraisals.

As can be seen in Figure 16 and Figure 17, for 1st line treatments sunitinib acts as a central node for all comparators of interest, with the exception of tivozanib. The networks are considerably more sparse for the risk subgroups (Appendix E) with no available risk subgroup

Kaplan-Meier curves for pembrolizumab + lenvatinib for PFS due to redaction in the NICE submission; in addition, OS subgroup data were not available for avelumab + axitinib. Risk subgroup Kaplan Meier curves were also not available for pazopanib for either OS of PFS. For the favourable risk subgroup the only trials of treatments recommended in this population where Kaplan Meier curves were available were CheckMate 9ER and JAVELIN Renal 101 and OS data was not available for JAVELIN Renal 101. Given this only time invariant NMA was conducted for the favourable risk subgroup. Proportional hazards NMAs at 2nd line-plus included all relevant comparators with the exception of pazopanib, as a reliable link could not be made to the network.

Figure 16: 1st line network diagram for PFS with summary HR and KM information

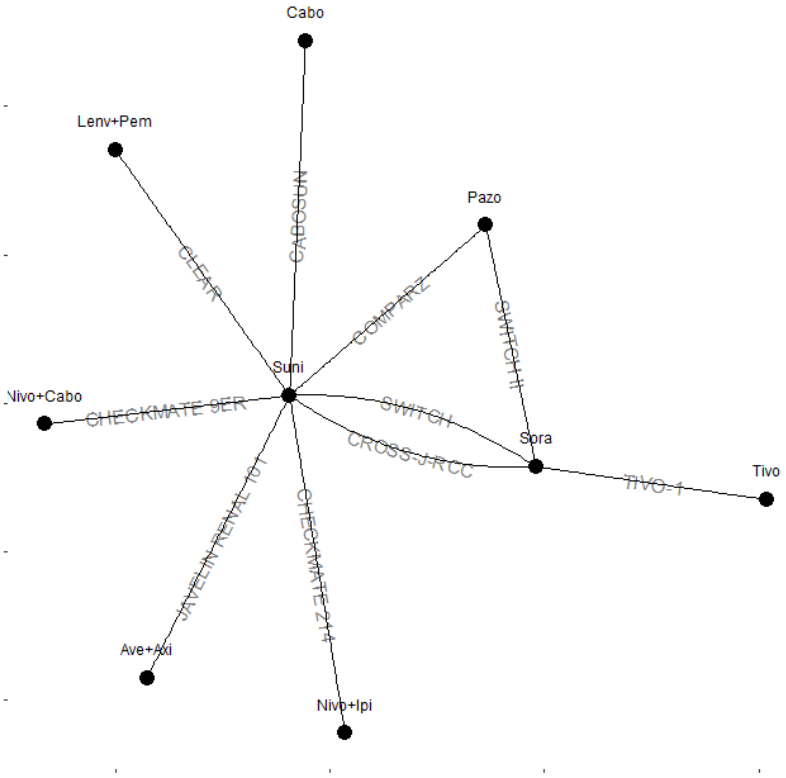
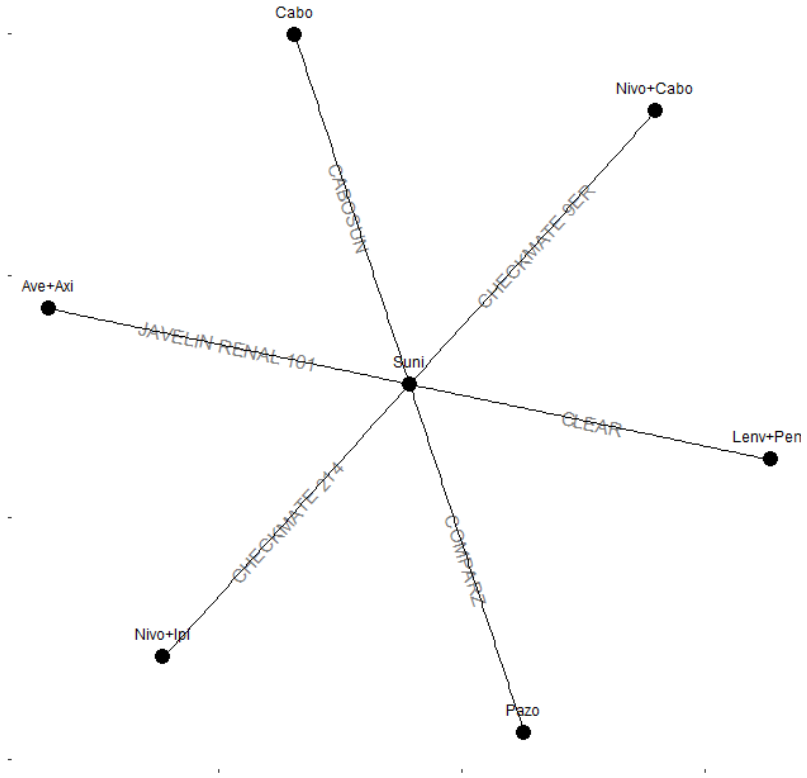


Figure 17: 1st line network diagram for OS with summary HR and KM information



3.7.2.1. Investigation of proportional hazards

Appendix E contains log cumulative hazard plots for included trials. Results of tests for proportional hazards using Schoenfeld residuals (i.e. Grambsch-Therneau tests) and based on EAG's digitisation of curves are provided in Table 36. Because these tests are based on our digitisations, there are likely small differences between the EAG's tests and published results; however, we were unable to precisely replicate results from CheckMate 9ER despite having IPD, possibly due to not being able to include stratifying factors in the analysis. In sum, there was clear and consistent evidence of non-proportional hazards across the network and for both outcomes. This is including with respect to key trials in the analysis, including CheckMate 9ER (also discussed in Section 3.4.3).

The EAG scrutinised log cumulative hazard plots alongside tests of proportional hazards. For PFS, visual assessment of proportional hazards was on several occasions at odds with significance tests. Aside from BERAT, where the small sample size meant a significance test would be underpowered, log cumulative hazard plots for CROSS-J-RCC, JAVELIN RENAL 101, SWITCH and TIVO-1 showed clear crossing of curves, in most cases on multiple time points. Plots with significant tests and visual checks suggesting non-proportionality included CheckMate 025, CheckMate 214, CheckMate 9ER, CLEAR, METEOR, and TIVO-3. Patterns in plots for CheckMate 025, CheckMate 214, CLEAR and TIVO-3 suggested crossing of hazards as well as a change in patterns over the time horizon. For CheckMate 025, CheckMate 214, and TIVO-3, hazards diverged over time, whereas for CLEAR hazards come closer together over time. Patterns in the plot for CheckMate 9ER (which had marginal significance in the EAG's test) suggested a clear separation of hazards over time and for METEOR a coming together of hazards over time.

For OS, findings between visual inspection and statistical tests largely matched, with the exception of TIVO-1, where the two trial arms crossed during the analysis time. Other plots with non-significant tests did not have visually obvious violations of proportional hazards. Visual inspection of plots for CLEAR showed a clear crossing and coming back together, and for CheckMate 9ER a clear separation and coming back together at the end of the analysis time.

These results indicate that an assumption of proportional hazards is unlikely to be valid within either the 1st line or 2nd line-plus aRCC setting.

Table 36: Results of tests for proportional hazards in the all-risk group using Cox regression

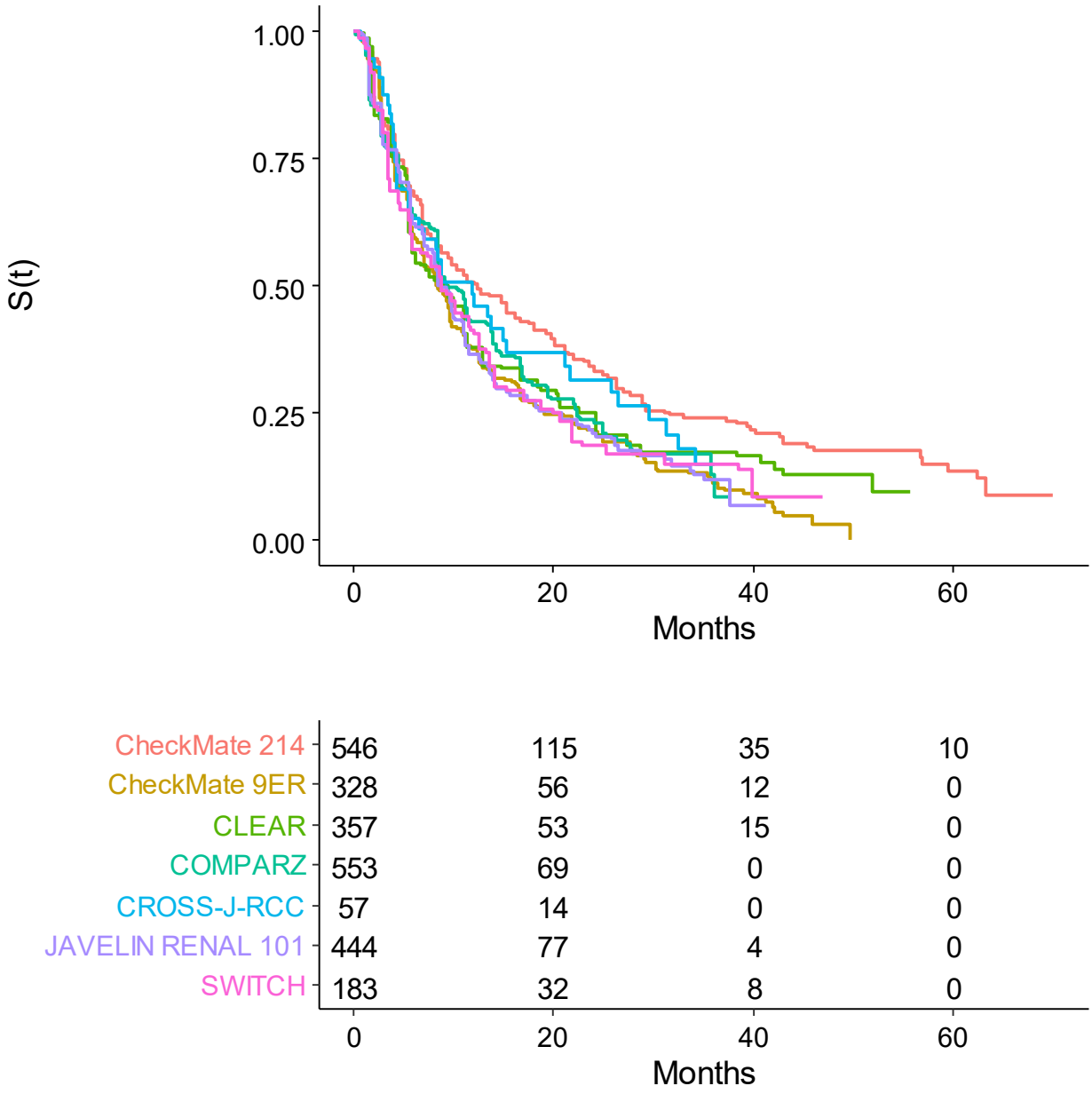
Study	P value: PFS	Visual check: PFS	P value: OS	Visual check: OS
AXIS	0.59	Yes	0.75	Yes
BERAT	0.13	No	NA	NA
CABOSUN	0.90	Yes	0.92	Yes
CheckMate 025	0.00016	No	0.34	Yes
CheckMate 214	0.000025	No	0.59	Yes
CheckMate 9ER	0.084	No	0.08	No
CLEAR	0.0027	No	0.00014	No
COMPARZ	0.25	Yes	0.44	Yes
CROSS-J-RCC	0.19	No	0.56	NA
JAVELIN RENAL 101	0.33	No	0.87	Yes
METEOR	0.032	No	0.56	Yes
NCT01136733	0.92	Yes	0.70	Yes
RECORD-1	0.66	Yes	0.31	Yes
SWITCH	0.15	No	0.32	NA
SWITCH II	0.72	Yes	0.43	NA
TIVO-1	0.29	No	0.83	No
TIVO-3	0.039	No	0.54	Yes

Note: Yes is no clear evidence of violation of proportional hazards; No represents evidence of violation of proportional hazards. Lenvatinib arm dropped from analysis for three-arm NCT01136733 trial

3.7.2.2. Effect modifiers across the network

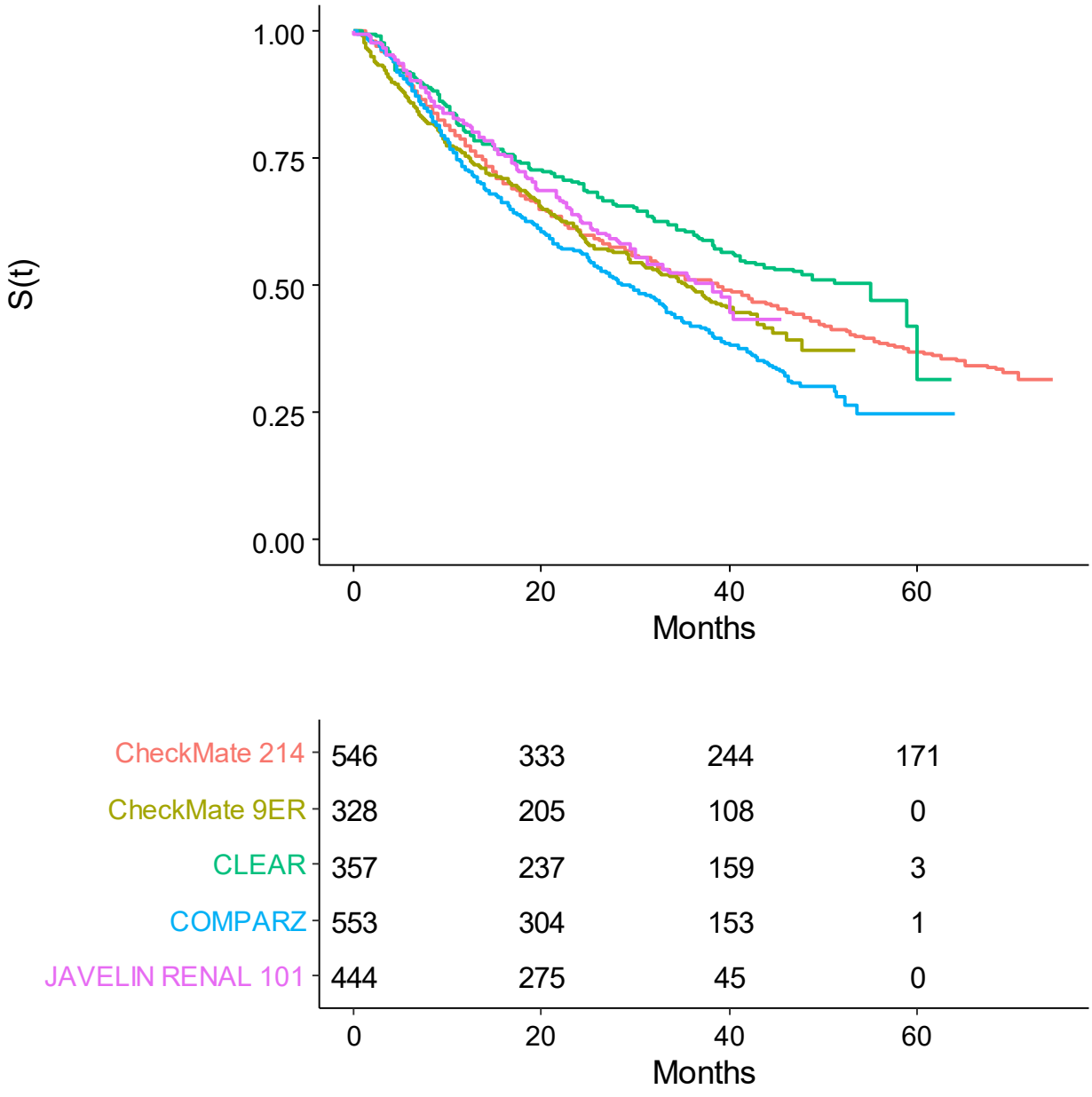
A central node within the network offers a common arm across the treatments which can be examined for heterogeneity in baseline risk. Survival data (PFS) for the sunitinib arms across the 1st line network are shown in Figure 18. Note that some digitisations were supplied at an earlier stage and may be updated with the final data-cut. There is some indication in the plot of anomalous PFS in the sunitinib arm of CheckMate214. There is no obvious explanation for this difference based on inclusion / exclusion criteria and baseline characteristics, and clinical experts consulted considered that this was most likely a chance observation; however, the EAG also noted that it could be due to use of investigator assessment for progression. For OS the COMPARZ trial looks to have anomalously low OS. This is to be expected as this trial was run prior to routine use of nivolumab as a subsequent therapy.

Figure 18: Survival data (PFS) for the central node (suni) of the 1st line network; all risk population



Abbreviations: PFS, progression free survival

Figure 19: Survival data (OS) for the central node (suni) of the 1st line network; all risk population

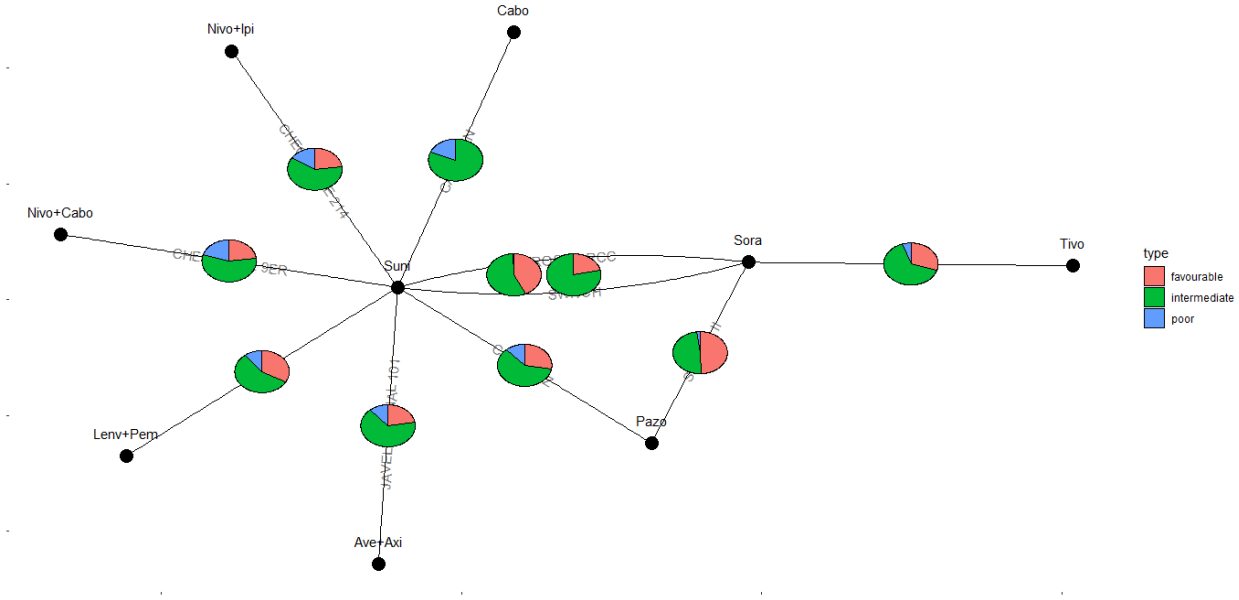


Abbreviations: OS, overall survival

Summary information for select potential effect modifiers is shown in Table 37. IMDC risk category is a primary effect modifier according to clinical advice.

A network graph for PFS of 1st line treatments overlaid with the proportions in risk subgroups is shown in Figure 20(following Cope et al[148]). This shows that the case mix is reasonably uniform across the network except for the three crossover trials that joined to the linking treatment sorafenib (which did not include poor risk patients) and the CABOSUN trial (which did not include favourable risk patients and is not recommended for use in this population). The expected impact of this is to bias towards tivozanib in the all risk population.

Figure 20: 1st line network with proportions of IMDC risk subgroups overlaid. The locations of pies are jittered when there are multiple trials between treatments



Notes: Three crossover trials (CROSS-J-RCC, SWITCH and SWITCH II) and one parallel group trial (TIVO-1) did not include (or included very few) poor risk patients, and the CABOSUN trial did not include favourable risk patients.

Abbreviations: Nivo: nivolumab; Ipi: ipilimumab; Cabo: cabozantinib; Sora: sorafenib; Tivo: tivozanib; Pazo: pazopanib; Ave: avelumab; Axi: axitinib; Lenv: lenvatinib; Pem: pembrolizumab

Table 37: Summary information for select effect modifiers

Trial name	Age (median) *	Risk status (%)¤			Line		Bone metastases (%) *	% clear cell	% prior nephrectomy	% sarcomatoid features
		Favourable	Intermediate	Poor	1L	2L+				
AXIS	61 | 61	20	64	16	0	100	NR	100	91	NR
BERAT	55	Included patients with up to 2 risk factors, split between favourable and intermediate not reported		0	0	100	NR	NR	20	NR
CABOSUN	63	0	81	19	100	0	NR	100	74.5	NR
CheckMate 025	62	36	49	15	0	100	18	100	88	NR
CheckMate 214	62 | 62	23	61	16	100	0	20 | 22	100	81.2	13
CheckMate 9ER	62 | 61	23	57	20	100	0	NR	100	69.9	11.9
CLEAR	64 | 62 | 61	32	55	10	100	0	24 | 24 | 27	100	74.6	6.8
COMPARZ	61 | 62	27	59	11	100	0	NR	100	83.2	NR
CROSS-J-RCC	67 | 67 | 66	21.7	78.3	0	100	0	23 | 33	100	88.3	NR
JAVELIN RENAL 101	62 | 61	22	62	16	100	0	NR	100	81.6	12
METEOR	62 | 63	46	42	13	0	100	22	100	85	NR
NCT01136733	61	23	37	40	0	100	27	100	88	NR
RECORD-1	61	29	56	14	0	100	35	100	97	NR
SWITCH	65	42	55	0.5	100	0	15	87	92	NR
SWITCH II	68 | 68	49	48	2	100	0	20	87	99	NR
TIVO-1	59 | 59	30	65	5	80	20	23 | 20	100	100	NR
TIVO-3	62 | 63	21	61	18	0	100	NR	100	NR	NR

Abbreviations: NR, not reported

* where results were available by arm the figures are shown separated by a bar (|).

¤ In some cases these do not add up to 100% due to rounding and risk status not having been recorded for some patients

Appendix E (Figures 17 to 23) presents the balance of other treatment effect modifiers across the 1st line network. The COMPARZ which links pazopanib to sunitinib has a lower proportion of patients with two or more metastatic sites than other studies which is likely to bias towards pazopanib. The SWITCH II and TIVO-1 trials had a larger proportion of patients with a prior nephrectomy which is likely to bias towards pazopanib and tivozanib. The TIVO-1 required a prior nephrectomy within the enrolment criteria. The CABOSUN trial had a larger proportion of patients with bone metastases enrolled; cabozantinib was considered by one of the experts consulted to be particularly effective in patients with bone metastases which may result in bias towards cabozantinib. Otherwise, patient characteristics were relatively well balanced across trials; particularly for trials of more recent treatments. Finally, the trials linking pazopanib and tivozanib to the network have a much lower proportion of subsequent IO use (or none) which will bias against these treatments when considering OS.

3.7.3. Results of time dependent NMA

The following sections contain summary results from frequentist and Bayesian analyses for all risk population and intermediate / poor risk population for OS and PFS at line 1. For line 1 PFS all risk, as the primary outcome, more detailed results are provided. Results for line 2+ are presented in Appendix E.

As explained above, sunitinib plays a central role in the 1st line networks and was selected as the reference treatment, along with CheckMate 9ER as the reference study. For 2nd line-plus networks, everolimus was chosen as the reference treatment and CheckMate 025 the reference study due to this being the treatment for which the longest follow-up was available.

A summary of the models selected by the process described in Section 3.7.1 is given in Table 38. As a note, AIC and DIC values that are lower reflect better fit compared to model complexity or parsimony. Generally, differences in AIC or DIC of between 3 and 5 values are considered noteworthy; however, the EAG generally preferred random effects models where these were supported by visual inspection and by the estimability of chosen models.

Table 38: Summary of final selected models for each line/risk/outcome subgroup

Outcome	Line	Risk group	Type	AIC	DIC	Exponent 1	Exponent 2
OS	1L	All	RE	1465.27	1466.5	-0.5	0.0
OS	2L+	All	RE	672.60	670.1	0.0	1.0

Outcome	Line	Risk group	Type	AIC	DIC	Exponent 1	Exponent 2
OS	1L	Intermediate/poor	FE	1121.26	1121.7	-0.5	0.5
PFS	1L	All	RE	1963.97	1982.0	-2.0	-0.5
PFS	2L+	All	RE	456.97	458.1	-0.5	0.5
PFS	1L	Intermediate/poor	RE	758.79	771.6	-2.0	-0.5

Abbreviations: 1L, 1st line; 2L+, 2nd line-plus; DIC: deviance information criterion; FE: fixed effects; OS: overall survival, PFS: progression free survival, RE: random effects

3.7.3.1. 1st line PFS all risk

The results of the frequentist model selection for PFS (1st line trials) are summarised in Table 39, which shows AIC values by the two exponents of each fractional polynomial fit. The model with lowest AIC has fractional polynomial exponents -2 and -0.5. In this instance, no other models attained AIC values within five points of the minimum.

Table 39: AIC values for fractional polynomial fit, 1st line PFS all risk

	-2	-1	-0.5	0	0.5	1	2	3
-2	-	1975.59	**1963.967**	1969.283	1996.790	2042.744	2148.740	2230.164
-1	-	-	1970.920	1994.467	2034.664	2085.816	2187.087	2258.683
-0.5	-	-	-	2021.301	2065.343	2115.107	2204.298	2262.540
0	-	-	-	-	2101.485	2144.774	2212.510	2250.925
0.5	-	-	-	-	-	2169.499	2209.582	2227.224
1	-	-	-	-	-	-	2200.388	2203.931
2	-	-	-	-	-	-	-	2185.450
3	-	-	-	-	-	-	-	-

Abbreviations: AIC: Akaike's information criterion; PFS: progression free survival

Notes: Row and column names correspond to exponent values. The model with lowest AIC is in bold. In this instance all other models had ΔAIC>5.

The fitted log-hazards under the NMA with the best-fitting (by AIC) fractional polynomial model are shown by trial in Figure 21. The trials approach a relatively constant hazard after about 20 to 40 months in each case. In some trials (e.g. CheckMate 9ER) there is an initial increase in hazard that inflects within the first 12 months.

A comparison of Bayesian model fits by fixed and random effects is shown in Table 40. In this case the random effects model has lower DIC. Hazard ratios from fitting by frequentist and Bayesian (random effects) methods are shown in Figure 22. Results are qualitatively similar. Survival curves under the Bayesian approach are shown in Figure 23 and Figure 24.

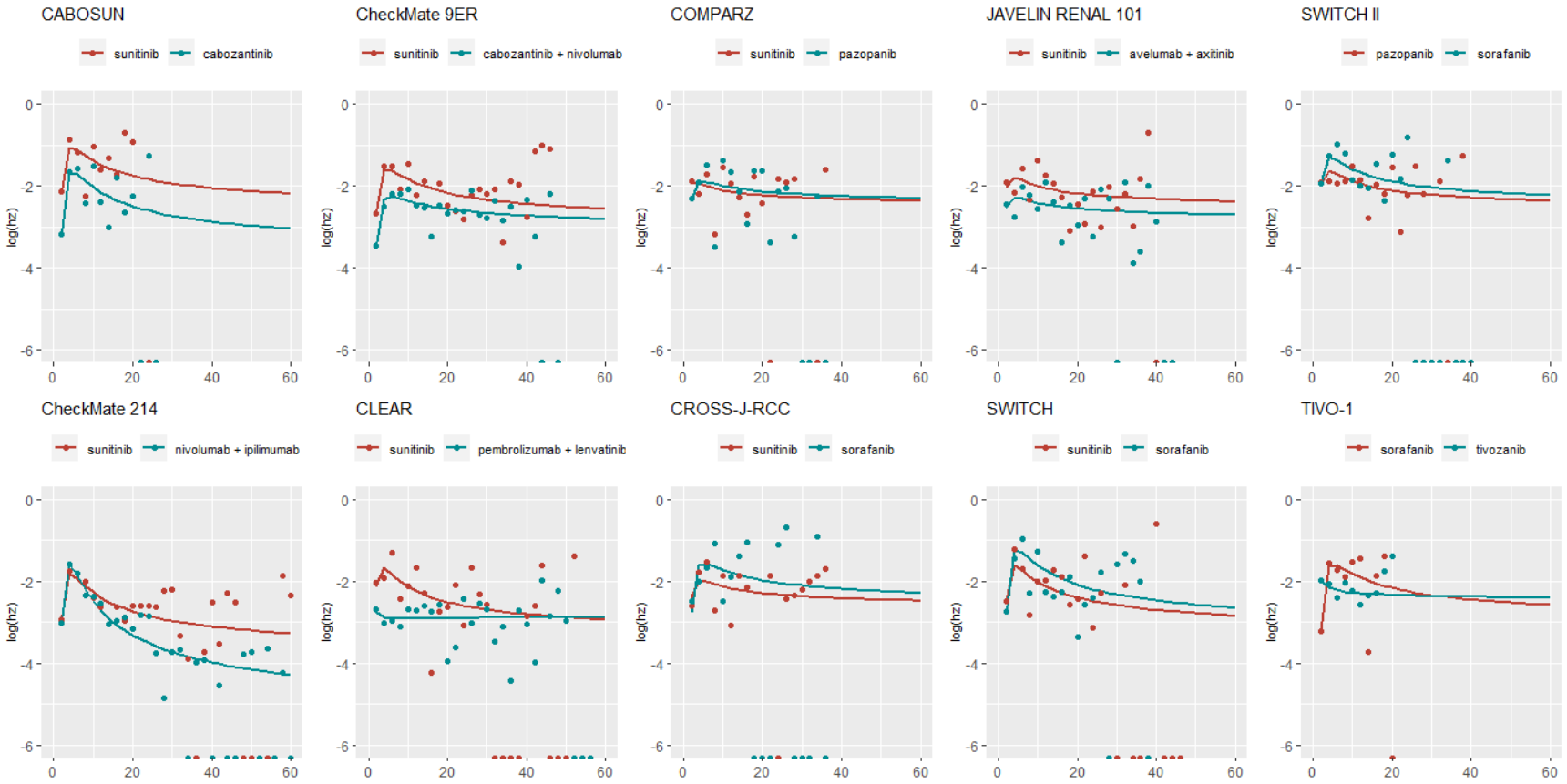
Table 40: Comparison of fixed and random effects Bayesian models for PFS for 1st line all risk

Model	Order	Exponents	DIC	pD	meanDev
FE	2	-2, -0.5	1983.2	53.9	1929.6
RE	2	-2, -0.5	1982	55	1927.1

Notes: using fractional polynomial model with exponents previously selected by frequentist methods.

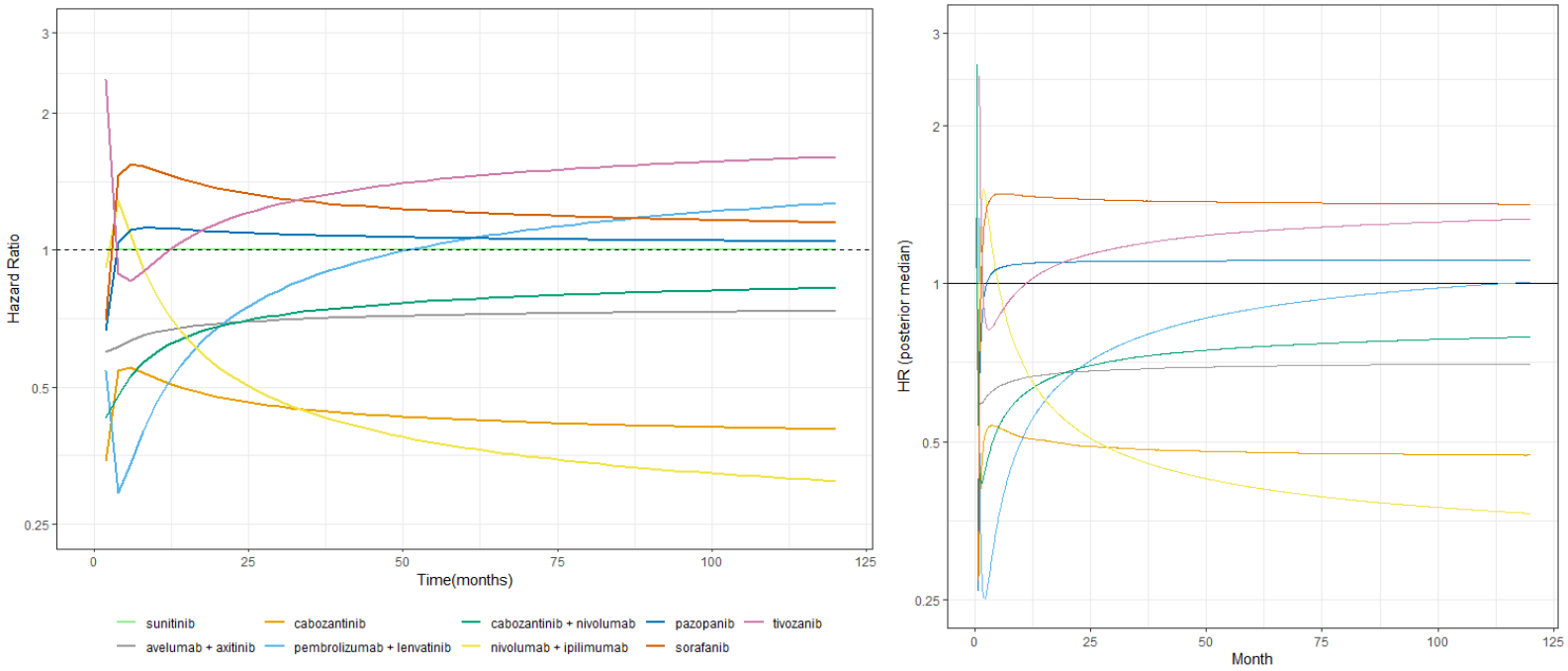
Abbreviations: DIC: Deviance Information Criteria, pD: effective number of parameters

Figure 21: Fitted log hazards for PFS for 1st line all risk



Notes: Fitted by fractional polynomial with exponents (-2, -0.5) across the network and extrapolated to 60 months. The points are logs of life-table estimates of the hazard (following Collett). Note that under sparse data the log hazard estimate is zero, which can be seen to the right of several plots as the event rate declines.

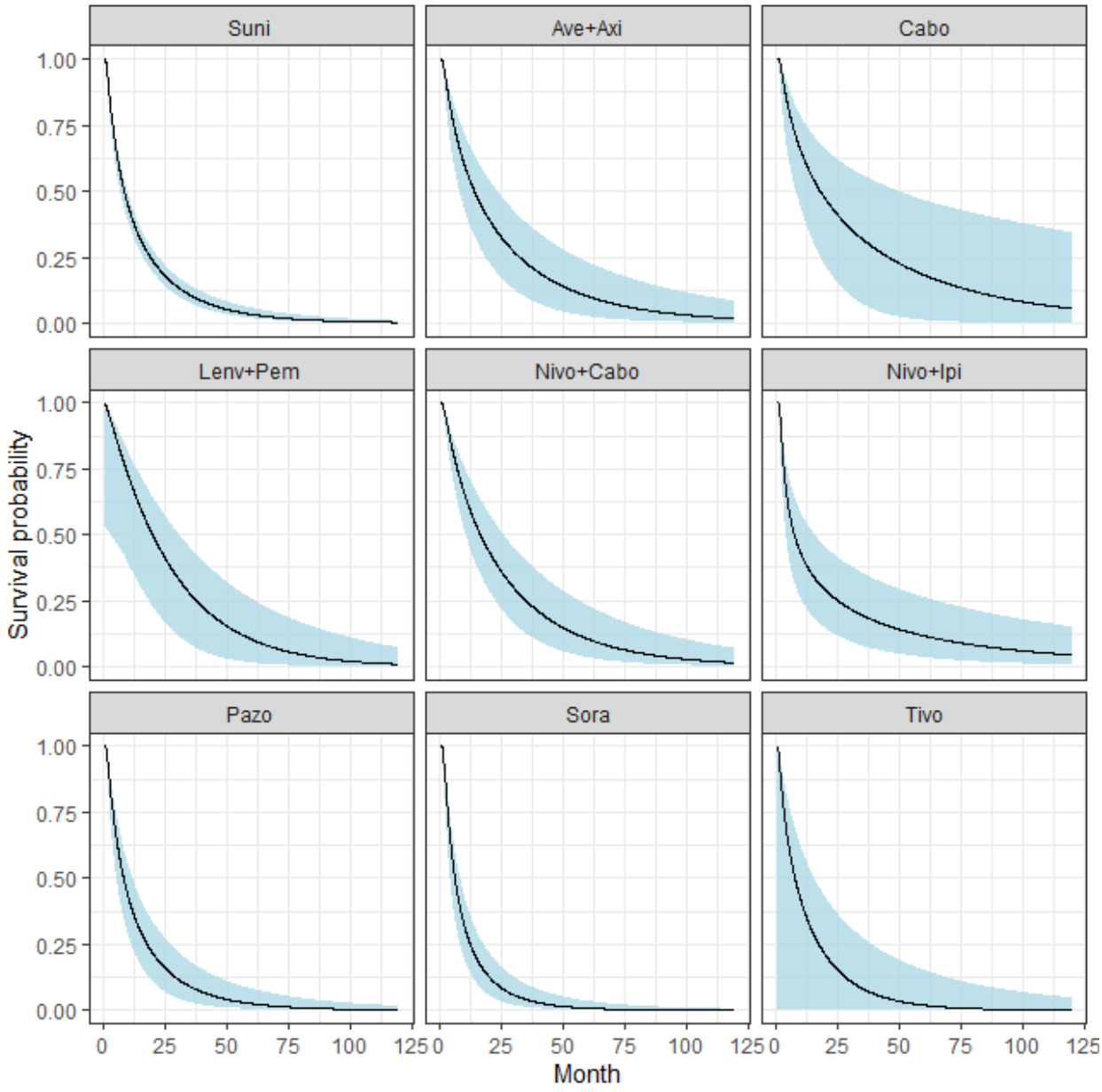
Figure 22: Time-dependent hazard ratios for PFS for 1st line all risk



Abbreviations: 1L, 1st line; PFS, progression free survival

Notes: Left : frequentist analysis. Right: Bayesian analysis (random effects). The reference treatment is sunitinib (central node in the network).

Figure 23: Survival curves by treatment from Bayesian fitting with the selected fractional polynomial model



Notes: Band is 95% credible interval.

A number of observations on the presented survival curves bear noting. First, HR plots in Figure 22 suggest that over time, treatments with higher HRs than sunitinib are other TKIs, whereas all other treatments than pembrolizumab + lenvatinib 'settle' into HRs less than 1 over the predicted time horizon. For cabozantinib + nivolumab the HR trends gradually upwards after the end of the observed data period, remaining below 1 during the first 60 months. Second, there is clear difference between treatments in the confidence bands surrounding fitted survival curves. This is perhaps most notable for cabozantinib and pembrolizumab + lenvatinib. For cabozantinib, this is likely due to the comparatively short timeframe included in analyses compared to other trials; whereas for pembrolizumab + lenvatinib, this may be due to comparatively poorer fit of the hazard function to the observed hazards in Figure 21. It should be noted that cabozantinib, nivolumab + ipilimumab, and pembrolizumab + lenvatinib are only recommended for intermediate and poor risk patients.

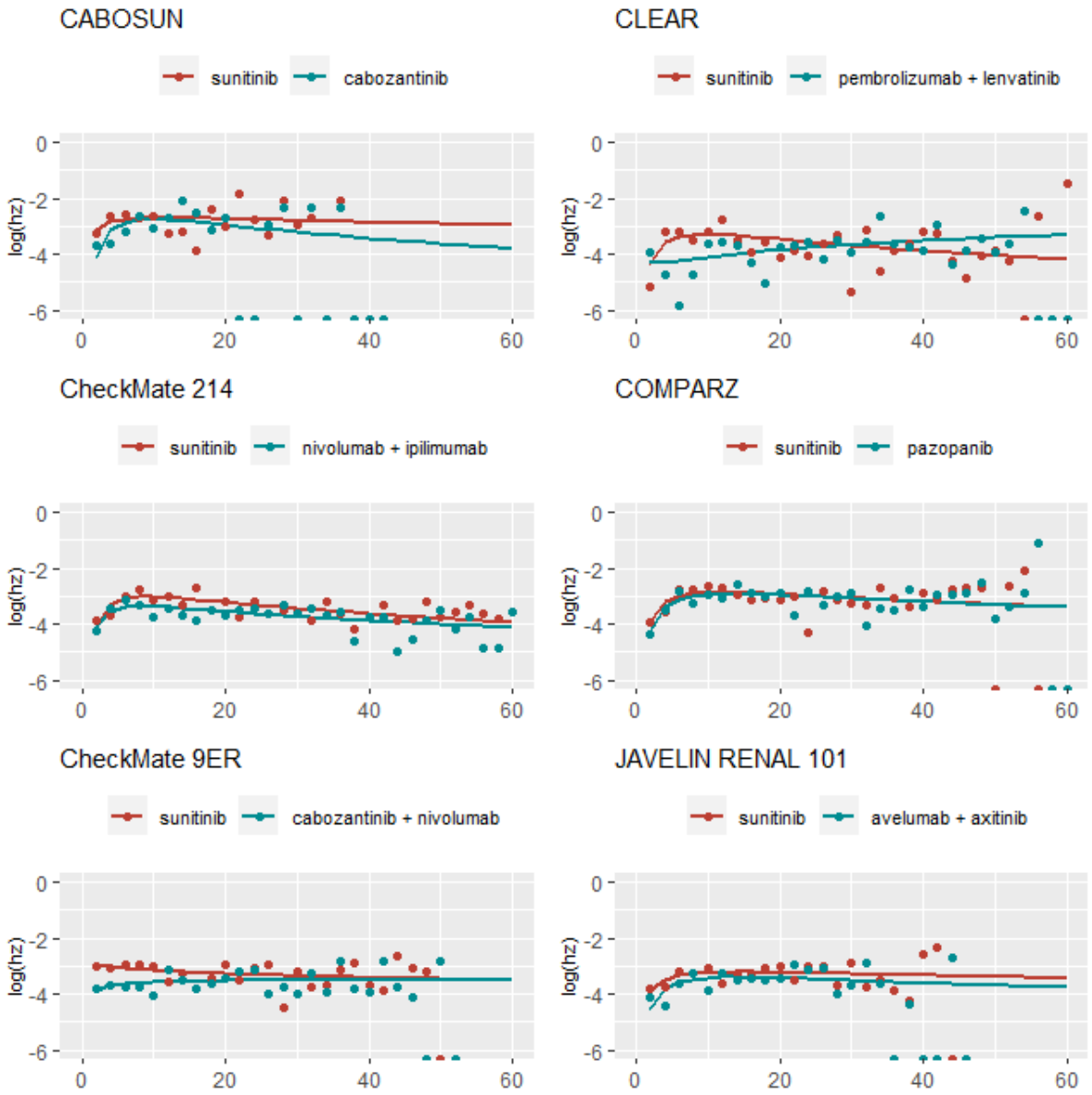
3.7.3.2. 1st line OS all risk

The selected model for first-line all risk OS had polynomial terms of -0.5 and 0. A number of models generated plausible AIC values; however, the chosen model had the best plausibility as assessed by the other criteria and based on input from expert elicitation. The very high initial HR for pembrolizumab + lenvatinib (

Figure 25) is associated with the unusual survival characteristics of the CLEAR trial, in which there were no or very few events in the sunitinib arm over the first two months (Motzer 2023).[149] The log-hazard (

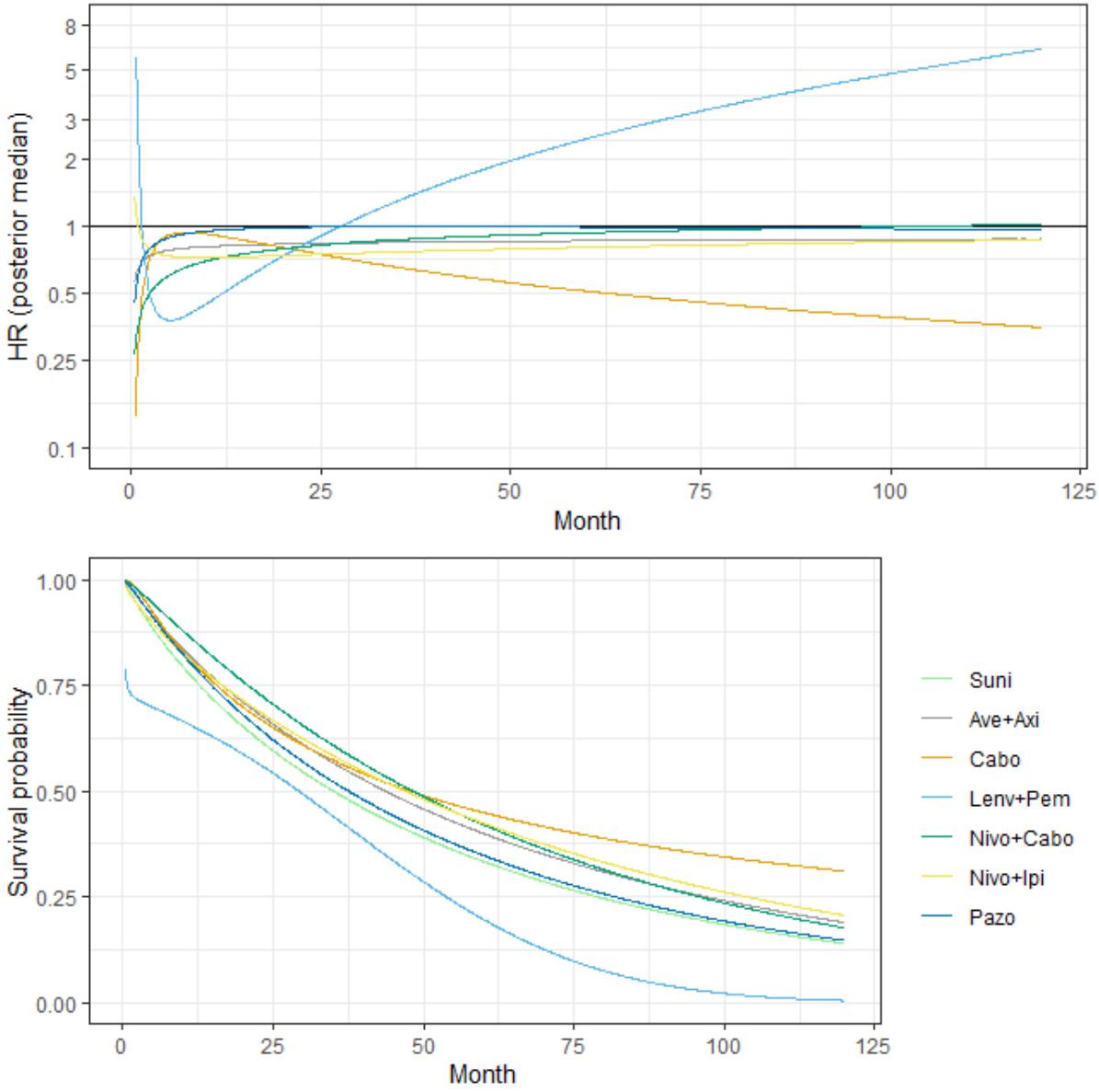
Figure 25) and survival curves (Figure 26) are qualitatively different to others in this subgroup, however, it should be noted that the expected survival for pembrolizumab + lenvatinib has high uncertainty, as can be seen in Figure 26. As with PFS in first line, cabozantinib has an unusually high level of uncertainty, likely due to the shorter timeframe of follow-up. Compared to PFS findings, findings for OS in this line are considerably more equivocal due possibly to the impact of subsequent treatments after progression; only cabozantinib appears to have a long-term HR substantially below 1 as compared to sunitinib. For cabozantinib + nivolumab again the HR trends gradually upwards after the end of the observed data period coming close to 1. There appears to be an early survival advantage for cabozantinib + nivolumab, especially relative to cabozantinib, that ends about month 50.

Figure 24: Log hazards for OS for 1st line all risk



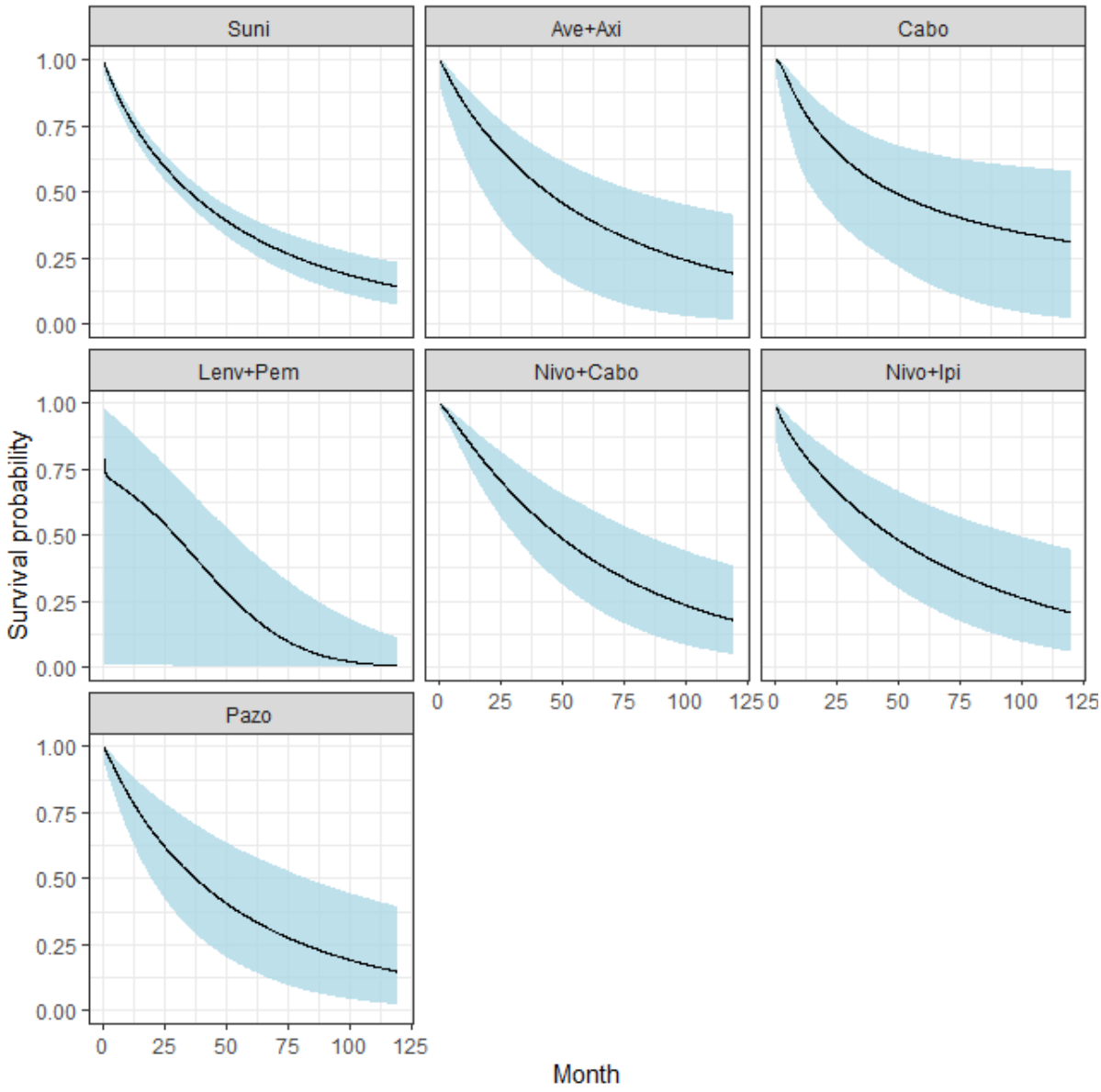
Abbreviations: 1L, 1st line; OS, overall survival

Figure 25: Hazard ratios and survival curves for OS for 1st line all risk (Bayesian analysis)



Abbreviations: Ave, avelumab; Axi, axitinib; Cabo, cabozantinib; Ipi, ipilimumab; Len, lenvatinib; Nivo, nivolumab; OS, overall survival; Pazo, pazopanib; Pem, pembrolizumab; Suni, sunitinib; vs, versus

Figure 26: Survival curves shown by treatment with 95% credible intervals.



Abbreviations: Ave, avelumab; Axi, axitinib; Cabo, cabozantinib; Ipi, ipilimumab; Len, lenvatinib; Nivo, nivolumab; Pazo, pazopanib; Pem, pembrolizumab; Suni, sunitinib

3.7.3.3. 1st line PFS intermediate/poor risk

Findings for PFS in first line for patients with intermediate or poor risk are presented in Figure 27 and Figure 28, with additional information in Appendix E (Figure 15, Table 11). The optimal model had polynomial terms of -2.0 and -0.5 and performed well in terms of AIC. The choice of model was also informed by expert elicitation, as estimates from these analyses better matched

the estimates from experts for novel therapies. We were unable to include pembrolizumab + lenvatinib in this analysis as Kaplan-Meier curves were not available for this subgroup. While all treatments show a long-term benefit in HRs as compared to sunitinib, these differences are unequal and highly uncertain for certain treatments. Time-varying HRs suggest that nivolumab with ipilimumab has a long-term lower HR than other treatments, reflected in a longer-term survival benefit emerging near the 60-month point. Cabozantinib monotherapy was predicted to have PFS similar to, or above, cabozantinib + nivolumab throughout the time period.

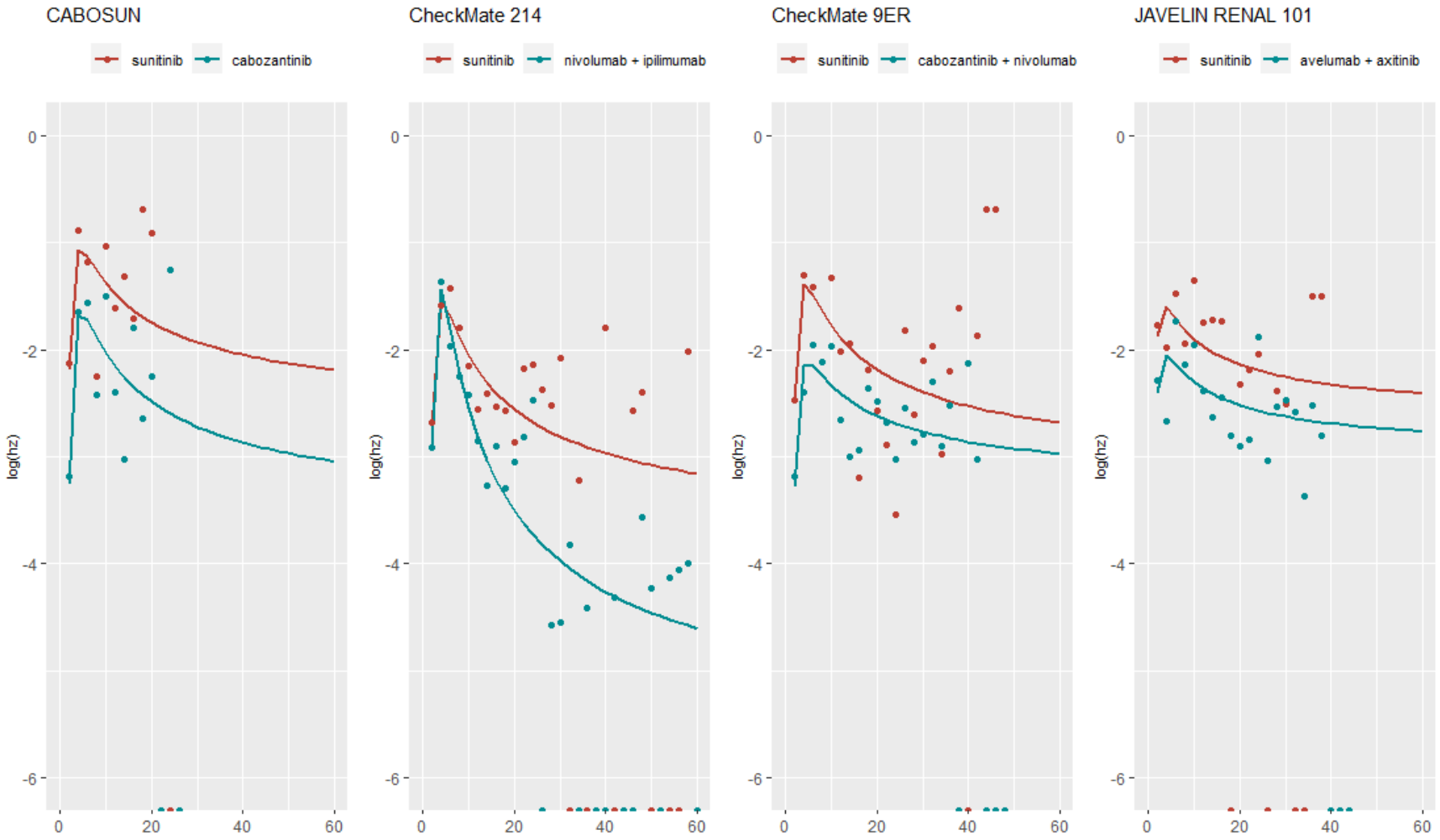
3.7.3.4. 1st line OS intermediate/poor risk

Findings for OS in first line for patients with intermediate or poor risk are presented in Figure 29 and Figure 30, with additional information in Appendix E (Figure 13, Table 12). The optimal model had polynomial terms of -0.5 and 0.5 and performed well relative to other models with AIC. Similar patterns of uncertainties in predicted survival curves were seen as in the analysis of PFS in intermediate and poor risk above. HR functions over time show a 'fanning out', with corresponding survival curves suggesting that different treatments have relatively better survival probabilities that change in order over the time horizon. Cabozantinib monotherapy was predicted to have OS similar to, or above, cabozantinib + nivolumab throughout the time period.

3.7.3.5. 2nd line-plus

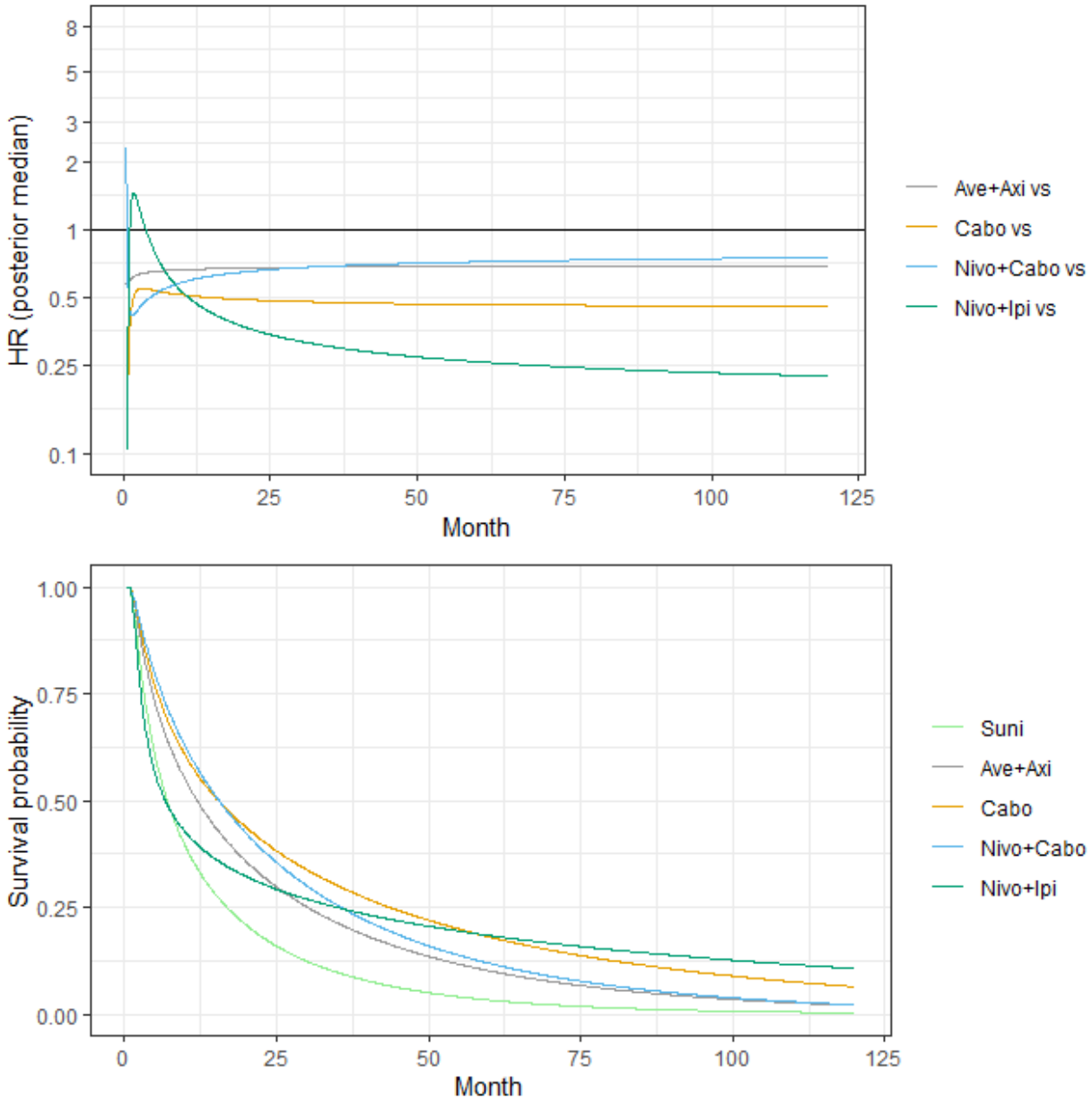
Findings for second-line and beyond outcomes are presented in Appendix E. We chose models that performed well in terms of AIC; furthermore, the PFS model was informed by expert elicitation to minimise the number of 0 or 1 probabilities in conditional survival at longer-term timepoints. Findings for PFS suggest a clear advantage in the survival function for lenvatinib plus everolimus until about 112 months, at which point it converges with nivolumab. Cabozantinib displays only limited improvement over everolimus which is unexpected given this is the 2nd line treatment favoured by clinicians. However, findings for OS suggest a different pattern, with cabozantinib possessing a long-term advantage in survival rates, followed by nivolumab. A contrasting misalignment was seen for everolimus plus lenvatinib, where PFS results were considerably more optimistic than OS results. In both situations, curves begin to display surprising results beyond the timepoints for which hazards where available, possibly due to the relatively limited follow-up time available from relevant trials to inform longer-term estimates. It should be stressed that predicted survival plots (Appendix E, Figure 14 and Figure 16) reflect substantial uncertainty.

Figure 27: Log hazards for PFS for 1st line intermediate/poor risk



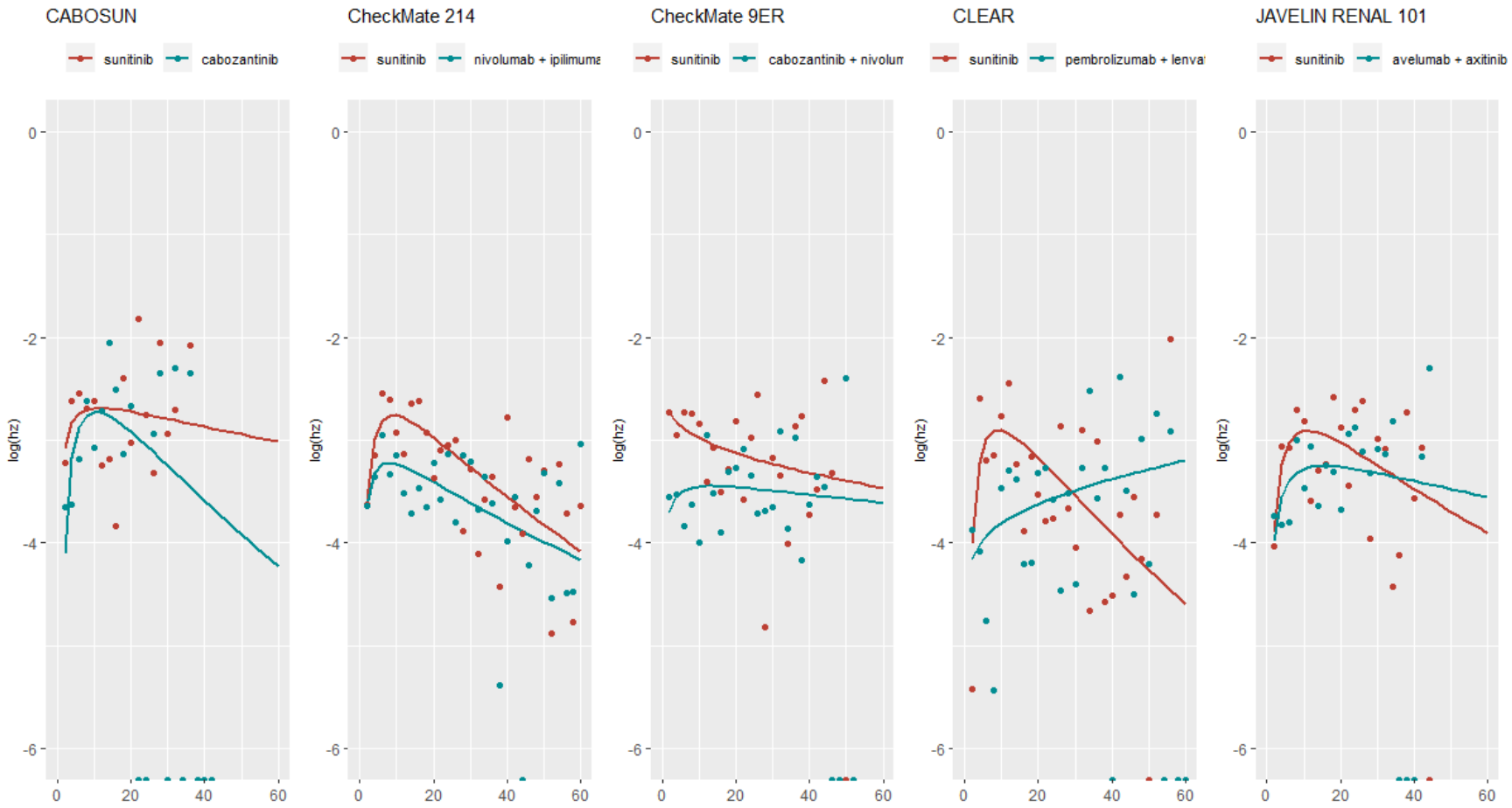
Abbreviations: PFS, progression free survival

Figure 28: Hazard ratios and survival curves for PFS for 1st line intermediate/poor risk (Bayesian analysis)



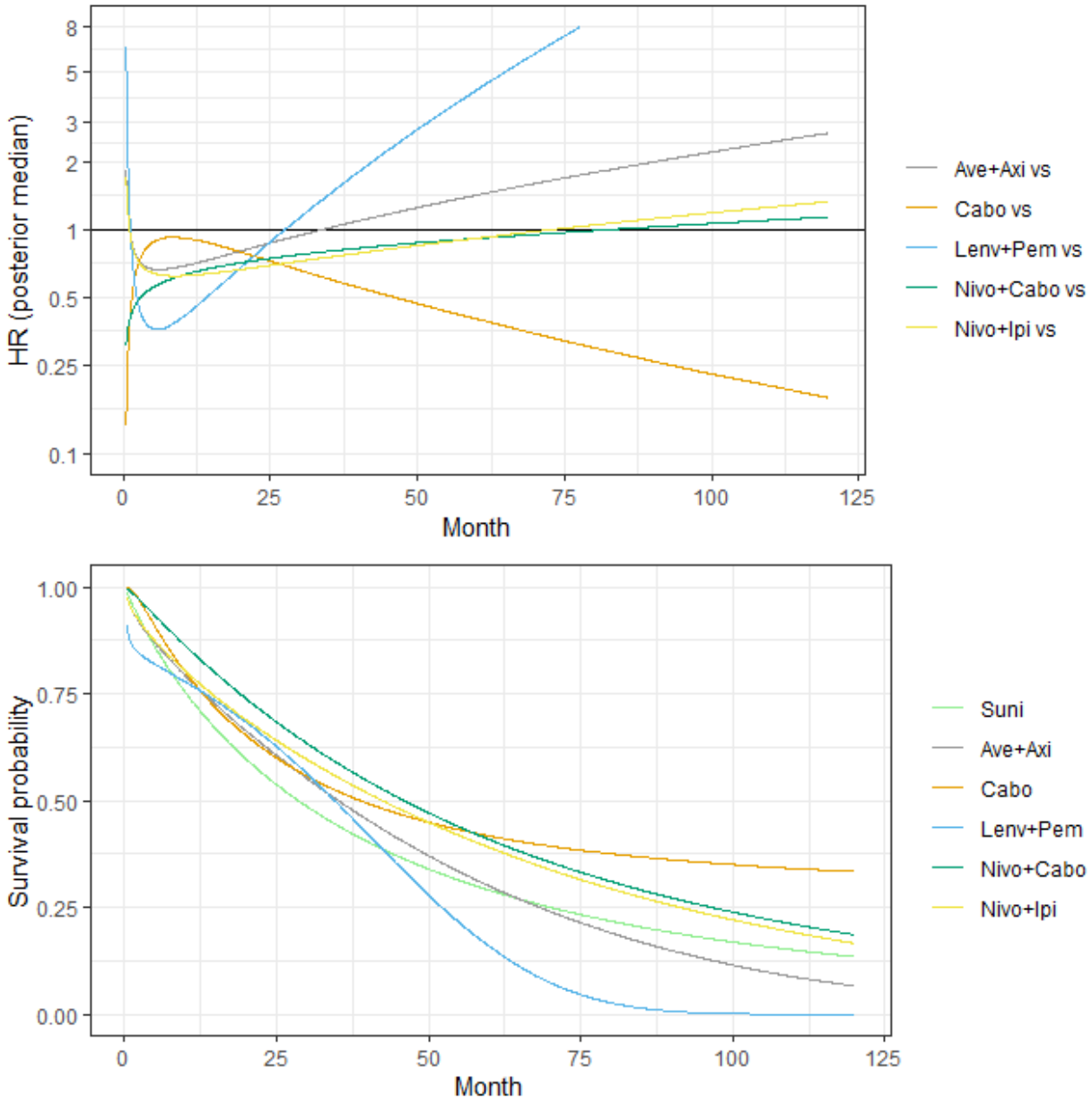
Abbreviations: Ave, avelumab; Axi, axitinib; Cabo, cabozantinib; Ipi, ipilimumab; Nivo, nivolumab; PFS, progression free survival; Suni, sunitinib

Figure 29: Log hazards for OS for 1st line intermediate/poor risk



Abbreviations: OS, overall survival

Figure 30: Hazard ratios and survival curves for OS for 1st line intermediate/poor risk (Bayesian analysis)



Abbreviations: Ave, avelumab; Axi, axitinib; Cabo, cabozantinib; Ipi, ipilimumab; Lenv, lenvatinib; Nivo, nivolumab; OS, overall survival; Pem, pembrolizumab; Suni, sunitinib

3.7.3.6. Interpretation and limitations

The EAG's fractional polynomial NMAs sought to compare different treatments in each network on the basis of time-varying HRs; i.e. by constructing the estimated HR for each treatment against a common comparator as a function of time. Using a multi-pronged assessment process, the EAG was able to select appropriate and justifiable models for each evidence network analysed. Importantly, the evidence of non-proportional hazards in a range of included trials (see Section 3.7.2.1) justified preference for a fractional polynomial method over a method assuming proportional hazards (i.e. inverse variance NMA using log HRs).

The EAG's analysis has a number of strengths. First, the use of a frequentist model selection stage followed by Bayesian analysis [143] of a subset meant it was practical for a large number of models to be efficiently assessed. At the frequentist model selection stage, all 2nd order fractional polynomial models (except repeated powers) were considered, creating 28 models per evidence network. At the Bayesian 'confirmatory' stage. a subset of models was used and compared for estimability and appropriateness, including a comparison of fixed effects and random effects (albeit time invariant). When random effects models were preferred by DIC, these generally offered only marginal improvement due to the large number of star networks analysed. However, in this analysis paradigm, time-invariant heterogeneity captured some of the difference between trials in common comparator hazards.

The EAG elected not to present a fractional polynomial NMA for the favourable risk group. This was justified on the basis of sparse availability of relevant Kaplan-Meier curves to support this analysis. Additionally, sparseness in networks, particularly in second-line plus, precluded inclusion of all relevant treatments; for example, axitinib could not be included in second-line and beyond. Moreover, differences in effect modifiers across network could cause bias in NMA. While the EAG did judge that NMAs were feasible, there was some broad variation over the network in effect modifiers identified through consultation, particularly in risk distribution. The CABOSUN trial was included in the 'all risk' population despite enrolling only intermediate/poor risk patients and the recommendation for cabozantinib being in the intermediate/poor risk population because the EAG did not regard that the difference between risk distributions was substantial enough to warrant its removal' however, it is notable as well that several trials did not enrol any poor risk patients. Uneven distributions of subsequent treatments may also have impacted interpretation of OS analyses in ways that are difficult to quantify across the network.

Finally, fractional polynomial NMAs require choice of model. While in some cases (particularly first-line all-risk PFS), AIC values clearly indicated the optimal model, in other cases AIC was not dispositive, and other sources of information were needed to determine optimal model choice. While expert elicitation for PFS outcomes was helpful, particularly at the five-year timepoint, it did not resolve all uncertainties in situations of multiple relevant choices. Thus, in the cost-effectiveness model, scenario analyses using proportional hazards NMAs are used as well.

3.7.4. Results of the time invariant NMA

We undertook NMAs for PFS, OS, ORR, discontinuation due to adverse events and risk of adverse events of grade 3 or higher. Adverse events data were only available in the ITT population. We present results for NMAs of the 1st line ITT population first, before presenting results for PFS, OS and ORR for intermediate/poor and favourable risk groups.

We interpreted the ITT population to be an 'all-comers' population and thus included all trials regardless of baseline risk distribution. This means, for example, that the CABOSUN trial was included despite only enrolling patients with intermediate or poor risk. We sensitivity analysed this assumption for the PFS outcome. Where we describe relevant treatments, we refer to those that are not included for linking (i.e. sorafenib) or for completeness (i.e. avelumab + axitinib). Finally, though all meta-analyses were undertaken in a Bayesian framework, we refer colloquially to 'statistical significance' where credible intervals do not include the point of unity.

3.7.4.1. Progression-free survival in 1st line ITT population

Base case analysis

Our proportional hazards NMA of PFS in the 1st line ITT population included all 10 relevant identified trials with 1st line groups. Because of the limited opportunities for estimation of heterogeneity in this NMA (one closed loop and only one comparison with more than one trial), we estimated this model as a fixed-effects analysis. Results are presented in Table 41 and suggested the numerical superiority of most relevant treatments against sunitinib except for pazopanib and tivozanib, but not a statistical difference of sunitinib against nivolumab + ipilimumab, pazopanib and tivozanib.

Cabozantinib + nivolumab was statistically better than nivolumab + ipilimumab, pazopanib, sunitinib and tivozanib, and was numerically, but not statistically, less effective than

cabozantinib alone and pembrolizumab + lenvatinib. However, it should be acknowledged that CABOSUN, the trial for cabozantinib alone vs sunitinib enrolled only intermediate or poor risk patients, for which the magnitude of treatment effects tends to be larger. Moreover, the CABOSUN trial used a higher dose of cabozantinib than other trials including this drug, which clinical advice suggests is linked to higher effectiveness in a dose-response relationship.

Because of the limited number of studies per comparison, we were unable to undertake network meta-regression to explore differences by study in key characteristics. However, we undertook two sensitivity analyses by assessor and presence of a poor-risk population.

We compared our model against an unrestricted mean effects model to evaluate the presence of inconsistency. The deviance information criterion (DIC) for our consistency model was 18.37, with a total residual deviance of 10.40. In contrast, the DIC for our unrestricted mean effects model was 18.74, with a total residual deviance of 9.72. This suggested that the consistency model was acceptable. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Preferring investigator-assessed PFS instead of blinded review PFS

Where PFS was presented at the latest datacut with both investigator-assessed and blinded independent central review, we preferred blinded review-based PFS. However, two trials (CABOSUN, COMPARZ) presented PFS at last datacut assessed via both methods. We used a fixed-effects analysis and found that results were very similar to the base case analysis (see Table 42).

We compared our model against an unrestricted mean effects model to evaluate the presence of inconsistency. The deviance information criterion (DIC) for our consistency model was 17.75, with a total residual deviance of 9.78. In contrast, the DIC for our unrestricted mean effects model was 18.58, with a total residual deviance of 9.64. This suggested that the consistency model was acceptable. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Excluding trials that did not enrol patients with poor risk

Three trials in our network (SWITCH, SWITCH II and CROSS-J-RCC) excluded patients with poor risk. We thus excluded these trials in a sensitivity analysis. The impact of this was to cause TIVO-1, and thus tivozanib, to be dropped from the network as all connecting trials evaluating

sorafenib were excluded. Results from this analysis are presented in Table 43. Findings for included treatments were very similar to the base case analysis.

No consistency results were generated as there were no closed loops in this network. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Table 41: PFS in 1st line ITT population (base case)

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+axi	-	1.136 (0.888,1.46)	1.405 (0.879,2.216)	0.78 (0.619,0.981)	0.668 (0.54,0.825)	1.425 (1.099,1.845)	0.491 (0.387,0.62)	0.671 (0.57,0.789)	0.65 (0.46,0.924)
Cabo+nivo	0.88 (0.685,1.126)	-	1.237 (0.765,1.98)	0.687 (0.538,0.882)	0.588 (0.467,0.742)	1.254 (0.948,1.646)	0.432 (0.336,0.557)	0.591 (0.49,0.711)	0.571 (0.401,0.825)
Cabo	0.712 (0.451,1.137)	0.809 (0.505,1.308)	-	0.556 (0.352,0.882)	0.476 (0.304,0.755)	1.012 (0.632,1.658)	0.349 (0.22,0.56)	0.478 (0.311,0.739)	0.462 (0.27,0.793)
Nivo+ipi	1.283 (1.019,1.615)	1.456 (1.134,1.859)	1.8 (1.134,2.839)	-	0.857 (0.693,1.053)	1.826 (1.411,2.364)	0.628 (0.497,0.794)	0.86 (0.732,1.009)	0.83 (0.586,1.185)
Pazo	1.496 (1.212,1.852)	1.701 (1.348,2.139)	2.101 (1.325,3.289)	1.167 (0.95,1.443)	-	2.134 (1.67,2.716)	0.734 (0.614,0.874)	1.005 (0.876,1.15)	0.974 (0.71,1.331)
Pem+lenv	0.702 (0.542,0.91)	0.797 (0.607,1.054)	0.989 (0.603,1.583)	0.548 (0.423,0.709)	0.469 (0.368,0.599)	-	0.344 (0.265,0.45)	0.471 (0.387,0.577)	0.456 (0.315,0.665)
Sora	2.036 (1.613,2.583)	2.317 (1.796,2.979)	2.864 (1.785,4.553)	1.592 (1.259,2.013)	1.362 (1.144,1.628)	2.91 (2.223,3.773)	-	1.368 (1.153,1.62)	1.322 (1.014,1.72)
Suni	1.49 (1.268,1.755)	1.692 (1.407,2.042)	2.092 (1.354,3.213)	1.162 (0.991,1.365)	0.995 (0.87,1.141)	2.124 (1.733,2.587)	0.731 (0.617,0.867)	-	0.967 (0.709,1.321)
Tivo	1.538 (1.083,2.176)	1.75 (1.212,2.494)	2.165 (1.261,3.699)	1.205 (0.844,1.707)	1.027 (0.752,1.409)	2.195 (1.505,3.174)	0.756 (0.581,0.986)	1.034 (0.757,1.411)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the HR metric. The comparison is the row-forming treatment against the column-forming treatment.

Table 42: PFS in 1st line ITT population (using investigator-assessed outcome at latest datacut)

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+axi	-	1.134 (0.89,1.453)	1.198 (0.777,1.863)	0.78 (0.624,0.977)	0.692 (0.562,0.856)	1.427 (1.105,1.846)	0.498 (0.396,0.634)	0.67 (0.571,0.787)	0.66 (0.466,0.94)
Cabo+nivo	0.882 (0.688,1.124)	-	1.055 (0.676,1.66)	0.686 (0.537,0.879)	0.61 (0.485,0.765)	1.259 (0.96,1.655)	0.439 (0.341,0.563)	0.59 (0.49,0.709)	0.583 (0.402,0.838)
Cabo	0.835 (0.537,1.288)	0.948 (0.602,1.479)	-	0.65 (0.421,1.014)	0.577 (0.376,0.884)	1.194 (0.757,1.885)	0.417 (0.268,0.646)	0.559 (0.372,0.843)	0.551 (0.33,0.923)
Nivo+ipi	1.283 (1.023,1.602)	1.457 (1.138,1.862)	1.537 (0.986,2.376)	-	0.888 (0.72,1.093)	1.832 (1.416,2.37)	0.64 (0.505,0.809)	0.859 (0.732,1.011)	0.848 (0.592,1.215)
Pazo	1.446 (1.168,1.781)	1.639 (1.307,2.061)	1.733 (1.132,2.656)	1.126 (0.915,1.388)	-	2.061 (1.63,2.626)	0.72 (0.605,0.859)	0.969 (0.848,1.106)	0.953 (0.696,1.313)
Pem+lenv	0.701 (0.542,0.905)	0.794 (0.604,1.042)	0.837 (0.53,1.321)	0.546 (0.422,0.706)	0.485 (0.381,0.614)	-	0.35 (0.266,0.457)	0.47 (0.385,0.573)	0.465 (0.317,0.676)
Sora	2.007 (1.578,2.527)	2.275 (1.776,2.937)	2.4 (1.549,3.729)	1.563 (1.236,1.978)	1.389 (1.164,1.653)	2.857 (2.189,3.763)	-	1.343 (1.132,1.599)	1.324 (1.023,1.729)
Suni	1.493 (1.271,1.752)	1.695 (1.41,2.04)	1.788 (1.187,2.688)	1.164 (0.989,1.365)	1.032 (0.904,1.18)	2.13 (1.744,2.596)	0.745 (0.626,0.883)	-	0.988 (0.72,1.356)
Tivo	1.515 (1.063,2.147)	1.716 (1.193,2.485)	1.814 (1.084,3.028)	1.179 (0.823,1.689)	1.049 (0.762,1.438)	2.152 (1.48,3.16)	0.755 (0.578,0.977)	1.013 (0.738,1.39)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the HR metric. The comparison is the row-forming treatment against the column-forming treatment.

Table 43: PFS in 1st line ITT population (excluding trials with poor risk exclusion)

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Suni
Ave+axi	-	1.137 (0.889,1.453)	1.393 (0.874,2.191)	0.778 (0.62,0.983)	0.638 (0.511,0.797)	1.426 (1.104,1.849)	0.67 (0.571,0.787)
Cabo+nivo	0.88 (0.688,1.124)	-	1.227 (0.764,1.971)	0.685 (0.537,0.877)	0.561 (0.44,0.715)	1.256 (0.956,1.648)	0.59 (0.49,0.71)
Cabo	0.718 (0.456,1.144)	0.815 (0.507,1.308)	-	0.558 (0.352,0.881)	0.457 (0.29,0.726)	1.024 (0.633,1.655)	0.48 (0.313,0.743)
Nivo+ipi	1.285 (1.018,1.612)	1.459 (1.14,1.863)	1.791 (1.135,2.843)	-	0.82 (0.655,1.02)	1.831 (1.416,2.366)	0.86 (0.73,1.011)
Pazo	1.567 (1.255,1.956)	1.781 (1.398,2.271)	2.189 (1.378,3.452)	1.22 (0.981,1.527)	-	2.236 (1.733,2.885)	1.05 (0.897,1.225)
Pem+lenv	0.701 (0.541,0.906)	0.796 (0.607,1.046)	0.976 (0.604,1.58)	0.546 (0.423,0.706)	0.447 (0.347,0.577)	-	0.47 (0.385,0.574)
Suni	1.492 (1.27,1.752)	1.696 (1.409,2.04)	2.083 (1.347,3.194)	1.163 (0.989,1.37)	0.952 (0.816,1.115)	2.126 (1.743,2.6)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the HR metric. The comparison is the row-forming treatment against the column-forming treatment.

3.7.4.2. Overall survival in 1st line ITT population

Our proportional hazards NMA of OS in the 1st line ITT population included six relevant identified trials with 1st line groups. We excluded trials testing sequences of treatments (CROSS-J-RCC, SWITCH, SWITCH II) as the OS estimates from these trials test sequences instead of individual treatments. As a result, we also excluded TIVO-1, and thus tivozanib, as this was now disconnected from the network. We estimated this model as a fixed-effects analysis as only one trial was available for each direct comparison, and we did not explore inconsistency as there were no closed loops in the network. Results are presented in Table 44 and suggested the numerical superiority of all treatments against sunitinib, though not the statistical superiority of cabozantinib or pazopanib. Results also did not suggest the superiority of any treatment against any other, with the exception of nivolumab with ipilimumab against pazopanib, though the pattern of effects suggested that cabozantinib with nivolumab was numerically superior to all other relevant treatments. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

3.7.4.3. Overall response rate in 1st line ITT population

Our NMA of ORR in the 1st line ITT population included all 10 relevant identified trials with 1st line groups. Because of the limited opportunities for heterogeneity in this NMA (one closed loop and only one comparison with more than one trial), we estimated this model as a fixed-effects analysis. We included the whole-population estimate from TIVO-1 in order to ensure tivozanib was represented in the network, since line-specific estimates for ORR were not available for this trial. Results are presented in Table 45 and suggested the numerical superiority of all relevant treatments against sunitinib, but not the statistical superiority of tivozanib. Cabozantinib with nivolumab was statistically superior to nivolumab with ipilimumab, pazopanib, sunitinib and tivozanib, numerically but not statistically superior to cabozantinib, and numerically but not statistically less effective than pembrolizumab with lenvatinib.

We compared our model against an unrestricted mean effects model to evaluate the presence of inconsistency. The deviance information criterion (DIC) for our consistency model was 39.53, with a total residual deviance of 21.47. In contrast, the DIC for our unrestricted mean effects model was 39.35, with a total residual deviance of 20.39. This suggested that the consistency model was acceptable. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Table 44: OS in 1st line ITT population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Suni
Ave+axi	-	1.128 (0.833,1.518)	0.984 (0.623,1.581)	1.096 (0.844,1.422)	0.859 (0.669,1.103)	0.999 (0.734,1.355)	0.789 (0.644,0.97)
Cabo+nivo	0.887 (0.659,1.2)	-	0.875 (0.552,1.404)	0.973 (0.744,1.278)	0.762 (0.585,1.001)	0.889 (0.641,1.215)	0.7 (0.56,0.878)
Cabo	1.016 (0.632,1.605)	1.143 (0.712,1.813)	-	1.113 (0.713,1.74)	0.873 (0.558,1.357)	1.012 (0.635,1.628)	0.804 (0.529,1.214)
Nivo+ipi	0.912 (0.703,1.185)	1.028 (0.783,1.345)	0.898 (0.575,1.403)	-	0.784 (0.631,0.973)	0.913 (0.69,1.193)	0.72 (0.614,0.843)
Pazo	1.164 (0.907,1.494)	1.312 (0.999,1.708)	1.145 (0.737,1.791)	1.276 (1.028,1.584)	-	1.165 (0.885,1.522)	0.92 (0.792,1.063)
Pem+lenv	1.001 (0.738,1.363)	1.125 (0.823,1.559)	0.988 (0.614,1.575)	1.096 (0.838,1.449)	0.858 (0.657,1.13)	-	0.789 (0.632,0.995)
Suni	1.267 (1.031,1.554)	1.428 (1.14,1.785)	1.243 (0.824,1.889)	1.39 (1.186,1.628)	1.087 (0.941,1.262)	1.267 (1.005,1.582)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the HR metric. The comparison is the row-forming treatment against the column-forming treatment.

Table 45: Overall response rate in 1st line ITT population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+ axi	-	0.961 (0.632,1.47)	1.174 (0.401,3.063)	2.306 (1.598,3.358)	2.163 (1.495,3.108)	0.732 (0.485,1.12)	3.813 (2.507,5.755)	3.14 (2.39,4.154)	2.339 (1.317,4.101)
Cabo+ nivo	1.041 (0.68,1.581)	-	1.234 (0.412,3.232)	2.415 (1.604,3.62)	2.254 (1.497,3.415)	0.765 (0.484,1.205)	3.975 (2.509,6.297)	3.277 (2.383,4.546)	2.438 (1.333,4.437)
Cabo	0.852 (0.326,2.49 7)	0.81 (0.309,2.42 9)	-	1.965 (0.768,5.726)	1.834 (0.71,5.397)	0.624 (0.241,1.863)	3.231 (1.231,9.67)	2.666 (1.085,7.527)	1.993 (0.712,6.341)
Nivo+ip i	0.434 (0.298,0.62 6)	0.414 (0.276,0.62 3)	0.509 (0.175,1.302)	-	0.936 (0.667,1.308)	0.316 (0.212,0.472)	1.65 (1.101,2.456)	1.36 (1.07,1.733)	1.011 (0.577,1.761)
Pazo	0.462 (0.322,0.66 9)	0.444 (0.293,0.66 8)	0.545 (0.185,1.409)	1.068 (0.764,1.5)	-	0.339 (0.227,0.502)	1.763 (1.284,2.425)	1.454 (1.146,1.849)	1.082 (0.653,1.776)
Pem+ lenv	1.367 (0.893,2.06 3)	1.307 (0.83,2.066)	1.603 (0.537,4.151)	3.16 (2.119,4.714)	2.954 (1.993,4.401)	-	5.205 (3.34,8.162)	4.288 (3.135,5.881)	3.193 (1.752,5.833)
Sora	0.262 (0.174,0.39 9)	0.252 (0.159,0.39 9)	0.31 (0.103,0.812)	0.606 (0.407,0.908)	0.567 (0.412,0.779)	0.192 (0.123,0.299)	-	0.825 (0.604,1.129)	0.615 (0.416,0.902)
Suni	0.318 (0.241,0.41 8)	0.305 (0.22,0.42)	0.375 (0.133,0.922)	0.735 (0.577,0.935)	0.688 (0.541,0.872)	0.233 (0.17,0.319)	1.212 (0.886,1.656)	-	0.745 (0.447,1.224)
Tivo	0.428 (0.244,0.75 9)	0.41 (0.225,0.75)	0.502 (0.158,1.404)	0.989 (0.568,1.734)	0.924 (0.563,1.531)	0.313 (0.171,0.571)	1.627 (1.109,2.406)	1.343 (0.817,2.235)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the OR metric. The comparison is the row-forming treatment against the column-forming treatment.

3.7.4.4. Discontinuation due to adverse events in 1st line ITT population

Our NMA of discontinuation due to AEs in the 1st line ITT population included all 10 relevant identified trials with 1st line groups. A fixed-effects model suggested inconsistency, with DIC (47.86) and total residual deviance (29.78) both higher than the corresponding values for the unrestricted mean effects model (DIC 38.70, total residual deviance 19.66). We then considered a random effects model using a stabilising prior distribution from Turner (2015[144]), in the form of a lognormal distribution with parameters (-2.29, 1.58^2). The resultant model showed satisfactory consistency when compared to an unrestricted mean effects model with the same informative prior distribution in respect of both DIC (39.68 vs 39.32) and total residual deviance (20.29 vs 19.76). One possible reason for this inconsistency is that evidence on discontinuation due to adverse events is inconsistently reported across included trials. In four trials, we extracted data from PRISMA flowcharts describing discontinuations due to adverse events; in another five trials, we extracted data from the text describing withdrawals or any treatment-emergent adverse event leading to treatment stop. It is possible that these outcome definitions generated some methodological heterogeneity in our NMA for this outcome. In addition, we included the whole-population estimate from TIVO-1 in order to ensure tivozanib was represented in the network, since line-specific estimates for discontinuation were not available for this trial.

Results are presented in Table 46. Nearly all credible intervals embraced 1, without a clear pattern of effects across treatments; comparisons between relevant treatments that were not sunitinib did not identify any statistically meaningful pairwise differences. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

3.7.4.5. Risk of treatment-emergent adverse events of grade 3 of higher in 1st line ITT population

Our NMA of risk of TEAEs of grade 3 or higher in the 1st line ITT population included all 10 relevant identified trials with 1st line groups. Because of the limited opportunities for heterogeneity in this NMA (one closed loop and only one comparison with more than one trial), we estimated this model as a fixed-effects analysis. We included the whole-population estimate from TIVO-1 in order to ensure tivozanib was represented in the network, since line-specific estimates for grade 3 or higher adverse events were not available for this trial. Results are presented in Table 47 and suggested a diverse pattern of effects. Cabozantinib with nivolumab had a statistically greater odds of TEAEs of grade 3 or higher as compared to nivolumab with ipilimumab, pazopanib, sunitinib, and tivozanib; numerically but not statistically greater odds

than cabozantinib; and numerically but not statistically lower odds than pembrolizumab with lenvatinib.

We compared our model against an unrestricted mean effects model to evaluate the presence of inconsistency. The deviance information criterion (DIC) for our consistency model was 37.42, with a total residual deviance of 19.23. In contrast, the DIC for our unrestricted mean effects model was 39.04, with a total residual deviance of 20.03. This suggested that the consistency model was acceptable. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Table 46: Discontinuation due to adverse events in 1st line ITT population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+axi	-	1.048 (0.314,3.367)	2.435 (0.612,9.354)	1.104 (0.341,3.614)	2.735 (0.945,8.014)	0.672 (0.211, 2.111)	2.741 (0.992,7.901)	2.393 (1.034,5.554)	2.634 (0.646,11.195)
Cabo+nivo	0.954 (0.297,3.18)	-	2.311 (0.585,9.078)	1.058 (0.324,3.456)	2.597 (0.909,7.622)	0.641 (0.197, 2.101)	2.612 (0.953,7.556)	2.296 (0.981,5.285)	2.513 (0.618,10.548)
Cabo	0.411 (0.107,1.635)	0.433 (0.11,1.709)	-	0.457 (0.117,1.76)	1.129 (0.331,3.994)	0.276 (0.07, 1.065)	1.138 (0.343,3.953)	0.989 (0.34,2.876)	1.085 (0.237,5.247)
Nivo+ipi	0.906 (0.277,2.929)	0.945 (0.289,3.083)	2.187 (0.568,8.516)	-	2.471 (0.838,7.282)	0.603 (0.192, 1.902)	2.489 (0.869,7.122)	2.166 (0.924,4.954)	2.414 (0.557,10.007)
Pazo	0.366 (0.125,1.058)	0.385 (0.131,1.1)	0.886 (0.25,3.023)	0.405 (0.137,1.193)	-	0.244 (0.085, 0.692)	1.009 (0.513,1.97)	0.881 (0.443,1.676)	0.975 (0.283,3.181)
Pem+lenv	1.488 (0.474, 4.732)	1.56 (0.476, 5.07)	3.617 (0.939, 14.26)	1.659 (0.526, 5.203)	4.091 (1.445, 11.829)	-	4.114 (1.46, 11.855)	3.564 (1.599, 8.196)	3.935 (0.871, 17.186)
Sora	0.365 (0.127,1.008)	0.383 (0.132,1.049)	0.879 (0.253,2.914)	0.402 (0.14,1.15)	0.991 (0.508,1.949)	0.243 (0.084, 0.685)	-	0.873 (0.463,1.586)	0.961 (0.348,2.645)
Suni	0.418 (0.18,0.967)	0.436 (0.189,1.019)	1.011 (0.348,2.943)	0.462 (0.202,1.082)	1.134 (0.597,2.256)	0.281 (0.122, 0.625)	1.145 (0.631,2.16)	-	1.1 (0.349,3.487)
Tivo	0.38 (0.089,1.547)	0.398 (0.095,1.618)	0.922 (0.191,4.215)	0.414 (0.1,1.797)	1.026 (0.314,3.534)	0.254 (0.058, 1.148)	1.041 (0.378,2.875)	0.909 (0.287,2.863)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the OR metric. The comparison is the row-forming treatment against the column-forming treatment.

Table 47: Risk of adverse events of grade 3 or higher in 1st line ITT population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+axi	-	0.702 (0.425,1.149)	0.891 (0.43,1.857)	1.818 (1.194,2.772)	1.114 (0.743,1.655)	0.576 (0.358,0.935)	1.348 (0.864,2.102)	1.197 (0.867,1.658)	1.966 (1.113,3.563)
Cabo+nivo	1.425 (0.87,2.353)	-	1.275 (0.594,2.714)	2.593 (1.641,4.121)	1.589 (1.01,2.509)	0.82 (0.49,1.385)	1.93 (1.182,3.132)	1.71 (1.167,2.518)	2.808 (1.531,5.179)
Cabo	1.122 (0.539,2.325)	0.784 (0.368,1.684)	-	2.042 (1.006,4.146)	1.252 (0.615,2.526)	0.646 (0.306,1.371)	1.516 (0.721,3.108)	1.342 (0.688,2.592)	2.205 (0.971,4.996)
Nivo+ipi	0.55 (0.361,0.838)	0.386 (0.243,0.609)	0.49 (0.241,0.994)	-	0.614 (0.427,0.878)	0.317 (0.205,0.491)	0.742 (0.497,1.106)	0.66 (0.501,0.86)	1.086 (0.631,1.88)
Pazo	0.898 (0.604,1.346)	0.629 (0.399,0.99)	0.799 (0.396,1.625)	1.628 (1.138,2.34)	-	0.518 (0.337,0.79)	1.209 (0.887,1.665)	1.076 (0.845,1.37)	1.769 (1.096,2.864)
Pem+lenv	1.736 (1.069,2.791)	1.22 (0.722,2.04)	1.548 (0.729,3.263)	3.156 (2.036,4.884)	1.93 (1.266,2.965)	-	2.342 (1.482,3.695)	2.078 (1.466,2.943)	3.425 (1.909,6.184)
Sora	0.742 (0.476,1.158)	0.518 (0.319,0.846)	0.66 (0.322,1.386)	1.347 (0.904,2.012)	0.827 (0.601,1.127)	0.427 (0.271,0.675)	-	0.887 (0.656,1.198)	1.462 (1.009,2.114)
Suni	0.836 (0.603,1.153)	0.585 (0.397,0.857)	0.745 (0.386,1.453)	1.514 (1.163,1.997)	0.929 (0.73,1.183)	0.481 (0.34,0.682)	1.128 (0.835,1.526)	-	1.646 (1.029,2.659)
Tivo	0.509 (0.281,0.899)	0.356 (0.193,0.653)	0.453 (0.2,1.03)	0.921 (0.532,1.586)	0.565 (0.349,0.912)	0.292 (0.162,0.524)	0.684 (0.473,0.991)	0.608 (0.376,0.972)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Notes: Findings are in the OR metric. The comparison is the row-forming treatment against the column-forming treatment.

3.7.4.6. Progression-free survival in 1st line intermediate or poor risk population

Our proportional hazards NMA of PFS in the 1st line intermediate or poor risk population included findings from nine trials (all 1st line trials except for SWITCH II). We included the estimate from TIVO-1 of PFS in the intermediate or poor risk population spanning 1st and 2nd line patients to ensure that tivozanib was represented in the network; otherwise, all estimates drew from 1st line patients only. The resultant network did not have any closed loops, and only the sunitinib-sorafenib comparison had more than one trial. Thus, we estimated a fixed-effects model. Results are presented in Table 48 and suggested all treatments were numerically superior to sunitinib, and statistically so for cabozantinib + nivolumab, cabozantinib, nivolumab + ipilimumab, and pembrolizumab + lenvatinib. Cabozantinib + nivolumab was statistically superior to pazopanib, sunitinib and tivozanib; numerically but not statistically superior to nivolumab + ipilimumab; and numerically but not statistically less effective than cabozantinib and pembrolizumab + lenvatinib. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

3.7.4.7. Overall survival in 1st line intermediate or poor risk population

Our proportional hazards NMA of OS in the 1st line intermediate or poor risk population included findings from six trials. Similar to the proportional hazards NMA of OS in the 1st line ITT population, we excluded CROSS-J-RCC and SWITCH. Findings from TIVO-1 and SWITCH II were not available for this outcome and risk group. The resultant network was star-shaped and no comparison had more than one trial in direct evidence. Thus, we estimated a fixed-effects model. Results are presented in Table 49 and suggested that all relevant treatments were superior to sunitinib. Cabozantinib + nivolumab was numerically superior to all relevant treatments, statistically so for pazopanib and sunitinib. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Table 48: PFS in 1st line intermediate/poor risk population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+axi	-	1.178 (0.904,1.542)	1.379 (0.863,2.176)	0.905 (0.704,1.168)	0.674 (0.517,0.879)	1.534 (1.142,2.062)	0.61 (0.426,0.87)	0.66 (0.552,0.789)	0.743 (0.479,1.146)
Cabo+nivo	0.849 (0.648,1.106)	-	1.168 (0.73,1.873)	0.767 (0.587,1.003)	0.572 (0.432,0.759)	1.303 (0.954,1.78)	0.516 (0.36,0.74)	0.561 (0.458,0.684)	0.629 (0.405,0.983)
Cabo	0.725 (0.46,1.159)	0.856 (0.534,1.369)	-	0.656 (0.414,1.034)	0.488 (0.305,0.778)	1.112 (0.691,1.815)	0.441 (0.263,0.747)	0.479 (0.313,0.735)	0.538 (0.299,0.963)
Nivo+ipi	1.105 (0.856,1.421)	1.304 (0.997,1.705)	1.525 (0.967,2.413)	-	0.746 (0.572,0.97)	1.699 (1.256,2.291)	0.672 (0.475,0.956)	0.729 (0.612,0.875)	0.82 (0.531,1.267)
Pazo	1.483 (1.137,1.935)	1.75 (1.318,2.316)	2.049 (1.285,3.284)	1.34 (1.031,1.75)	-	2.279 (1.682,3.098)	0.902 (0.629,1.308)	0.979 (0.803,1.198)	1.103 (0.706,1.731)
Pem+lenv	0.652 (0.485,0.876)	0.767 (0.562,1.049)	0.899 (0.551,1.448)	0.588 (0.437,0.796)	0.439 (0.323,0.595)	-	0.397 (0.269,0.579)	0.43 (0.339,0.547)	0.485 (0.301,0.765)
Sora	1.639 (1.149,2.345)	1.937 (1.352,2.779)	2.265 (1.34,3.804)	1.488 (1.046,2.106)	1.109 (0.764,1.59)	2.518 (1.728,3.719)	-	1.084 (0.803,1.469)	1.218 (0.946,1.58)
Suni	1.516 (1.267,1.81)	1.782 (1.463,2.186)	2.089 (1.361,3.195)	1.372 (1.143,1.635)	1.022 (0.834,1.245)	2.326 (1.829,2.946)	0.923 (0.681,1.245)	-	1.125 (0.755,1.674)
Tivo	1.345 (0.872,2.088)	1.59 (1.018,2.471)	1.858 (1.039,3.343)	1.22 (0.79,1.883)	0.907 (0.578,1.417)	2.063 (1.307,3.317)	0.821 (0.633,1.057)	0.889 (0.597,1.324)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

Table 49: OS in 1st line intermediate/poor risk population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pazo	Pem+lenv	Suni
Ave+axi	-	1.218 (0.877,1.669)	0.989 (0.622,1.578)	1.161 (0.885,1.526)	0.887 (0.67,1.179)	1.067 (0.761,1.495)	0.791 (0.636,0.982)
Cabo+nivo	0.821 (0.599,1.14)	-	0.814 (0.507,1.313)	0.958 (0.706,1.29)	0.73 (0.536,0.989)	0.882 (0.618,1.25)	0.651 (0.509,0.832)
Cabo	1.011 (0.634,1.608)	1.229 (0.762,1.974)	-	1.176 (0.759,1.832)	0.897 (0.579,1.399)	1.08 (0.671,1.746)	0.799 (0.533,1.206)
Nivo+ipi	0.861 (0.655,1.13)	1.044 (0.775,1.416)	0.851 (0.546,1.318)	-	0.763 (0.601,0.976)	0.92 (0.679,1.252)	0.68 (0.578,0.807)
Pazo	1.128 (0.848,1.492)	1.37 (1.011,1.864)	1.115 (0.715,1.727)	1.311 (1.024,1.663)	-	1.204 (0.887,1.637)	0.892 (0.749,1.061)
Pem+lenv	0.937 (0.669,1.315)	1.134 (0.8,1.619)	0.926 (0.573,1.491)	1.086 (0.799,1.474)	0.83 (0.611,1.128)	-	0.74 (0.574,0.959)
Suni	1.264 (1.019,1.572)	1.536 (1.201,1.965)	1.252 (0.829,1.876)	1.471 (1.24,1.731)	1.121 (0.942,1.336)	1.351 (1.043,1.743)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

3.7.4.8. Overall response rate in 1st line intermediate or poor risk population

Our NMA of ORR in the 1st line ITT population included findings from five trials (CABOSUN, CheckMate 214, CLEAR, JAVELIN Renal 101, CheckMate 9ER) for which data were available for this risk group, line and outcome. The resultant network was star-shaped and no comparison had more than one trial in direct evidence. Thus, we estimated a fixed-effects model. Results are presented in Appendix E and suggested that all treatments were superior to sunitinib. Cabozantinib + nivolumab was statistically superior to nivolumab + ipilimumab and sunitinib; numerically superior to cabozantinib; and statistically less effective than pembrolizumab + lenvatinib. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Table 50: Overall response rate in 1st line intermediate/poor risk population

	Ave+axi	Cabo+nivo	Cabo	Nivo+ipi	Pem+lenv	Suni
Ave+axi	-	0.851 (0.514,1.39)	1.197 (0.414,3.181)	1.609 (1.042,2.467)	0.485 (0.29,0.807)	3.17 (2.323,4.375)
Cabo+nivo	1.174 (0.72,1.946)	-	1.407 (0.482,3.862)	1.891 (1.175,3.047)	0.572 (0.325,0.994)	3.726 (2.542,5.518)
Cabo	0.835 (0.314,2.415)	0.711 (0.259,2.074)	-	1.347 (0.518,3.841)	0.406 (0.147,1.18)	2.662 (1.051,7.23)
Nivo+ipi	0.622 (0.405,0.96)	0.529 (0.328,0.851)	0.742 (0.26,1.93)	-	0.302 (0.184,0.49)	1.972 (1.483,2.636)
Pem+lenv	2.061 (1.24,3.449)	1.747 (1.006,3.079)	2.461 (0.848,6.783)	3.315 (2.04,5.442)	-	6.535 (4.418,9.821)
Suni	0.316 (0.229,0.43)	0.268 (0.181,0.393)	0.376 (0.138,0.951)	0.507 (0.379,0.674)	0.153 (0.102,0.226)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

3.7.4.9. PFS in 1st line favourable risk population

Our proportional hazards NMA of PFS in the 1st line favourable risk population included findings from eight of the nine trials that enrolled favourable risk patients (i.e. excluding SWITCH II). We included the estimate from TIVO-1 of PFS in the favourable risk population spanning 1st and 2nd line patients to ensure that tivozanib was represented in the network; otherwise, all estimates drew from 1st line patients only. The resultant network did not have any closed loops, and only the sunitinib-sorafenib comparison had more than one trial. Thus, we estimated a fixed-effects model. Results are presented in Table 51 and did not suggest a consistent pattern of effectiveness relative to sunitinib. Cabozantinib + nivolumab was numerically superior to all relevant treatments except for pembrolizumab + lenvatinib, and was statistically superior to

nivolumab + ipilimumab. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

3.7.4.10. OS in 1st line favourable risk population

Our proportional hazards NMA of OS in the 1st line favourable risk population included findings from five of the nine trials that enrolled favourable risk patients. Estimates were not available for TIVO-1, thus excluding tivozanib from the network, and we excluded both crossover trials for which estimates were available for this outcome (CROSS-J-RCC, SWITCH). The resultant network was star-shaped with one trial per comparison. Thus, we estimated a fixed-effects model. Results are presented in Table 52 and did not suggest any evidence of effectiveness relative to sunitinib. Cabozantinib + nivolumab was numerically, but not statistically, less effective than all relevant treatments. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

3.7.4.11. Overall response rate in 1st line favourable risk population

Our NMA of ORR in the 1st line favourable risk population included findings from four trials (CheckMate 214, CLEAR, JAVELIN Renal 101, CheckMate 9ER). The resultant network was star-shaped with one trial per comparison. Thus, we estimated a fixed-effects model. Results are presented in Table 53 and suggested that all treatments except for nivolumab + ipilimumab generated higher ORR in this population as compared to sunitinib; in contrast, nivolumab + ipilimumab generated worse ORR in this population. Cabozantinib + nivolumab was statistically superior to nivolumab + ipilimumab and sunitinib, and numerically superior to pembrolizumab + lenvatinib. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Table 51: PFS in 1st line favourable risk population

	Ave+axi	Cabo+nivo	Nivo+ipi	Pazo	Pem+lenv	Sora	Suni	Tivo
Ave+axi	-	0.985 (0.591,1.662)	0.444 (0.267,0.732)	0.7 (0.441,1.121)	1.416 (0.856,2.328)	0.451 (0.255,0.799)	0.708 (0.496,1.025)	0.761 (0.37,1.586)
Cabo+nivo	1.015 (0.602,1.692)	-	0.45 (0.265,0.741)	0.711 (0.434,1.144)	1.435 (0.854,2.386)	0.458 (0.255,0.817)	0.721 (0.489,1.051)	0.774 (0.369,1.6)
Nivo+ipi	2.254 (1.366,3.739)	2.222 (1.35,3.779)	-	1.58 (0.988,2.518)	3.2 (1.954,5.192)	1.024 (0.585,1.744)	1.6 (1.135,2.244)	1.733 (0.836,3.497)
Pazo	1.428 (0.892,2.27)	1.406 (0.874,2.306)	0.633 (0.397,1.012)	-	2.026 (1.256,3.238)	0.644 (0.38,1.1)	1.013 (0.744,1.373)	1.091 (0.539,2.178)
Pem+lenv	0.706 (0.43,1.168)	0.697 (0.419,1.17)	0.313 (0.193,0.512)	0.494 (0.309,0.796)	-	0.318 (0.181,0.56)	0.501 (0.35,0.715)	0.539 (0.262,1.102)
Sora	2.217 (1.252,3.919)	2.183 (1.224,3.928)	0.976 (0.573,1.709)	1.554 (0.909,2.634)	3.145 (1.786,5.516)	-	1.57 (1.021,2.422)	1.695 (1.076,2.624)
Suni	1.413 (0.976,2.015)	1.388 (0.952,2.045)	0.625 (0.446,0.881)	0.987 (0.728,1.345)	1.996 (1.399,2.861)	0.637 (0.413,0.979)	-	1.077 (0.572,1.997)
Tivo	1.313 (0.63,2.7)	1.293 (0.625,2.707)	0.577 (0.286,1.196)	0.917 (0.459,1.857)	1.856 (0.907,3.814)	0.59 (0.381,0.929)	0.929 (0.501,1.747)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pazo, pazopanib; pem, pembrolizumab; tivo, tivozanib

Table 52: OS in 1st line favourable risk population

	Ave+axi	Cabo+nivo	Nivo+ipi	Pazo	Pem+lenv	Suni
Ave+axi	-	0.612 (0.276,1.385)	0.699 (0.345,1.447)	0.748 (0.371,1.499)	0.704 (0.325,1.557)	0.66 (0.359,1.216)
Cabo+nivo	1.633 (0.722,3.626)	-	1.138 (0.593,2.16)	1.218 (0.654,2.244)	1.149 (0.551,2.294)	1.074 (0.635,1.786)
Nivo+ipi	1.43 (0.691,2.896)	0.879 (0.463,1.687)	-	1.068 (0.65,1.762)	1.002 (0.545,1.832)	0.944 (0.645,1.384)
Pazo	1.336 (0.667,2.696)	0.821 (0.446,1.53)	0.936 (0.568,1.538)	-	0.936 (0.532,1.671)	0.881 (0.634,1.223)
Pem+lenv	1.42 (0.642,3.078)	0.87 (0.436,1.814)	0.998 (0.546,1.836)	1.068 (0.598,1.881)	-	0.941 (0.583,1.513)
Suni	1.516 (0.822,2.786)	0.931 (0.56,1.576)	1.06 (0.722,1.549)	1.135 (0.818,1.576)	1.063 (0.661,1.716)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pazo, pazopanib; pem, pembrolizumab

Table 53: Overall response rate in 1st line favourable risk population

	Ave+axi	Cabo+nivo	Nivo+ipi	Pem+lenv	Suni
Ave+axi	-	1.494 (0.597,3.786)	9.113 (4.068,20.436)	1.767 (0.791,4.028)	3.695 (2.02,6.99)
Cabo+nivo	0.669 (0.264,1.674)	-	6.08 (2.656,14.196)	1.189 (0.509,2.815)	2.484 (1.277,4.97)
Nivo+ipi	0.11 (0.049,0.246)	0.164 (0.07,0.376)	-	0.195 (0.093,0.403)	0.407 (0.243,0.683)
Pem+lenv	0.566 (0.248,1.265)	0.841 (0.355,1.963)	5.139 (2.483,10.734)	-	2.085 (1.225,3.562)
Suni	0.271 (0.143,0.495)	0.403 (0.201,0.783)	2.456 (1.465,4.12)	0.48 (0.281,0.817)	-

Abbreviations: ave, avelumab; axi, axitinib; cabo, cabozantinib; ipi, ipilimumab; ITT, intention to treat; lenv, lenvatinib; nivo, nivolumab; pazo, pazopanib; pem, pembrolizumab; tivo, tivozanib

3.7.5. Cross-cutting commentary on network meta-analyses

Our time-invariant NMAs have a number of caveats in their interpretation, in addition to the comments offered in Section 3.7.3.6. First, time-invariant NMAs using summary effect sizes for survival outcomes (i.e. for OS and PFS outcomes) rely on an assumption of proportionality within comparisons entered into each model. This assumption was violated multiple times in our network as the assumption of proportional hazards was tenuous for at least one outcome in each included trial. While it is possible to interpret the HR from a model where the proportional hazards assumption has been violated as a time-average effect, it is likely preferable to use survival curves directly in indirect treatment comparisons. This was the basis for our fractional polynomial NMA. However, a competing issue that is posed by fractional polynomial NMAs is the need to undertake model selection. Like all extrapolation analyses, this introduces a degree of subjectivity to the analysis, but is likely to provide 'higher-fidelity' estimates of relative treatment effects.

Second, we used the most mature datacut available for each trial in all NMAs. This is a challenge for both fractional polynomial and time-invariant NMAs. While this is unlikely to have made a substantial difference for binary outcomes beyond a point of maturity, we are aware that there is some debate that equivalent timepoints should have been used across trials for analysis, generally because more mature data (for example, for overall survival) may reveal relationships not in evidence in earlier datacuts. We did not take this approach for several reasons. First, using earlier datacuts even where trials are highly mature would discard valuable information contributing to precision of effect sizes. Second, we did not regard that there was a good basis *ex ante* for grouping trial follow-up times, and it is likely that this would have led to the exclusion of trials reporting inadequately similar follow-up times. Third, while we did identify some evidence of maturing HRs over time, we did not identify consistent patterns in evolving shape of survival curves and trends in effect size when we jointly considered different levels of trial maturity and different treatments. In Figure 31 and Figure 32, we present examples from OS and PFS estimates in sequential datacuts for key trials. For three out of four IO/TKI combinations (i.e. excepting avelumab + axitinib), there appears to be slippage in OS estimates with sequential datacuts; the same trend is less in evidence for the one IO/IO combination (nivolumab + ipilimumab). Of interest is that the same trend in IO/TKI combinations is less immediately obvious for PFS outcomes. The mechanisms underpinning this evolution over time, and the mismatch in evolution, are unclear and merit further investigation.

Figure 31: Plot of cumulative OS over sequential datacuts in key trials

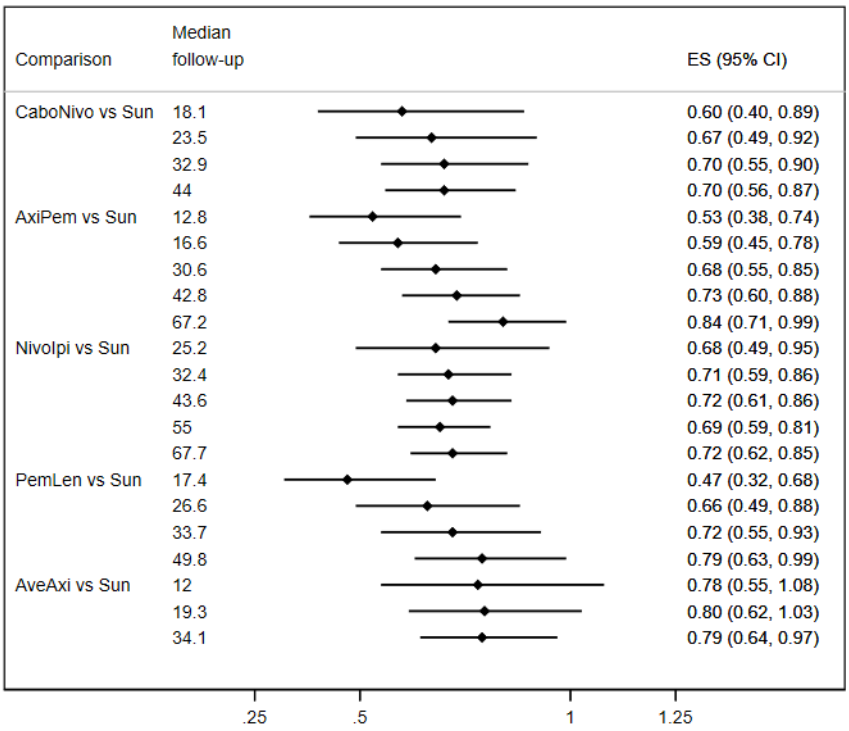
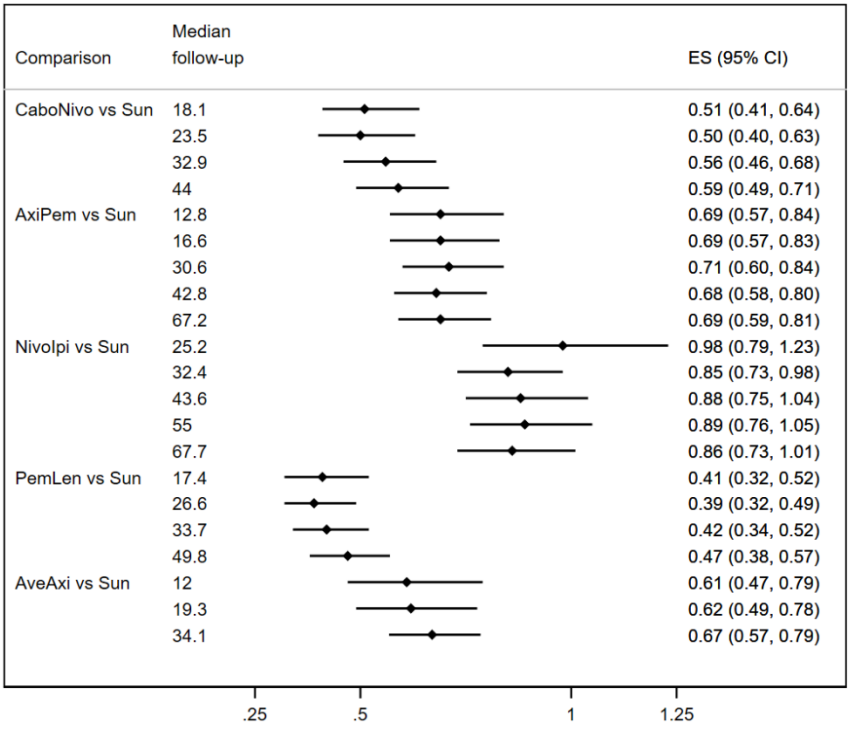


Figure 32: Plot of cumulative PFS over sequential datacuts in key trials



Third, most of our networks relied on one trial per direct comparison; even where networks had closed loops, these were sparse in the direct evidence available for each comparison. Again, this was a challenge for both fractional polynomial and time-invariant NMAs. The key limitation is that we were unable to account for differences over comparisons in the network in the distribution of potential effect modifiers.

Fourth, NMAs of safety outcomes are often unusually challenging given the diverse reporting of these outcomes. This is somewhat reflected in our findings relating to discontinuation due to AEs in the 1st line population. NMAs of safety outcomes should thus be regarded with some caution.

Finally, NMAs for 2nd line patient populations relied on a linking trial with a small sample size and documented issues with protocol administration. This mean that for some outcomes, networks were incomplete. These results should be interpreted in the view that not all relevant treatments were included in these meta-analyses.

3.7.6. Conclusions from the EAG NMAs

EAG NMAs included both fractional polynomial NMAs for OS and PFS and proportional hazards NMAs for the same outcomes, and NMAs for overall response rate and adverse event outcomes. On the whole, EAG NMAs reflected several challenges in this evidence base, including imbalanced distribution of effect modifiers, differences in follow-up, and challenges (particularly in second line) constructing evidence network, leading to the exclusion of tivozanib in some first line networks and axitinib and tivozanib in some second-line networks. However, both sets of NMAs reflect salient differences in effectiveness between treatments, particularly on PFS outcomes. As mentioned prior, inference on any differences in OS is complicated by subsequent treatment. A key issue comparing the NMAs was with respect to estimability in the CLEAR trial. The fractional polynomial NMA generated unreasonably pessimistic estimates of pembrolizumab + lenvatinib's effectiveness due to differences in events accumulated early in the time horizon, biasing results against this treatment; in contrast, the proportional hazards NMA provide an unduly favourable estimate of effectiveness given the convergence of hazards between treatment arms and the clear violations of the proportional hazards assumption.

Focusing on cabozantinib + nivolumab and the comparators relevant to the decision problem in each risk group, fractional polynomial NMAs for PFS and OS in the all risk group suggested that this combination was more effective than TKIs in first-line. Similarly, time-invariant NMAs for both OS and PFS reflected that cabozantinib + nivolumab was superior to TKIs.

In the intermediate/poor risk group, PFS for cabozantinib + nivolumab appeared to generate a predicted early survival benefit coterminous with cabozantinib up through about month 15, whereas for OS, cabozantinib + nivolumab generated an early survival advantage through 55 months, at which point survival curves with other treatments, including cabozantinib and nivolumab with ipilimumab, crossed. Time-invariant NMAs in the intermediate/poor risk population for both OS and PFS reflected that while cabozantinib + nivolumab was superior to TKIs, it was not generally statistically distinguishable from other novel treatments.

NMA estimates for the favourable risk group were only available from time-invariant NMAs. Cabozantinib + nivolumab was not generally distinguishable from other treatments in either OS or PFS analyses.

3.7.7. Comparison to other published network meta-analyses

To contextualise our findings, we compare our results to the ERG's NMAs in TA858, to the company's own NMAs, and to a recently published Cochrane review, all of which considered treatments in the 1st line. We also contrast our findings to the most recent NMA of 2nd line treatments in RCC, Liao 2022. Of note is that the only NMAs to also use a fractional polynomial method of those discussed below were presented by the company.

3.7.7.1. TA858

Our analysis strategy was similar to that undertaken in TA858, in that we undertook NMAs for PFS, OS, and ORR, and our analyses used a Bayesian framework. Similar to TA858, we preferred fixed effects models given the sparseness of included networks, and preferred blinded assessments of progression and response outcomes over investigator assessments. Unlike TA858, we interpreted the ITT population as an 'all-comers' group and thus included CABOSUN; we also considered all treatments for 1st line with available data in risk group analyses (e.g. even though sunitinib is not restricted by risk group, we included it in analyses of patients with intermediate or poor risk). We also used TIVO-1 to connect tivozanib to networks where estimates by line were unavailable, and were able to include JAVELIN-RENAL-101 and

CheckMate 9ER in our analyses. We further undertook NMAs of discontinuation due to AEs and risk of grade 3 or higher treatment-emergent adverse events as we were able to identify with appropriate reliability evidence from included RCTs. The ERG's concern in TA858 about the reliability of discontinuation evidence was somewhat reflected in the difficulties we faced with this analysis. Unlike TA858, we did not undertake sensitivity analysis by censoring rule used in PFS as we did not have the evidence available to undertake this. Specific comparison of results is limited both by differences in treatments included in each network and by the fact that we used different datacuts for several of the same trials used in TA858 (CheckMate 214, CLEAR).

3.7.7.2. Company-reported indirect treatment comparison

Our analysis strategy was also broadly similar to that undertaken in the company's NMA, including the use of a Bayesian framework. We used a proportional hazards NMA as an adjunct to additional methods for exploring differences between treatments in survival outcomes; however, we only considered fractional polynomial NMAs for survival outcomes as this was our protocol-specified method. We also did not consider proportion of intermediate or poor risk as a covariate due to the sparseness of the networks. A key difference between our NMAs and the NMAs reported by the company is that we included a systematically different set of trials. Unlike the company, we did not include KEYNOTE-426, which compared pembrolizumab with axitinib against sunitinib, as this was not a relevant comparator in this appraisal. We also did not consider the SUTENT trial (NCT01147822) separately from COMPARZ, as we used the pooled analyses from the COMPARZ studies in our NMAs. We also included data from the 1st line in crossover trials where these data were available (SWITCH, SWITCH II, CROSS-J-RCC), enabling use of TIVO-1 in several NMAs and thus including tivozanib in several networks. Where we undertook analyses of intermediate and poor risk patients, we pooled these groups as a result of our feasibility assessment. Finally, we included CABOSUN in our ITT population.

For fractional polynomial NMAs, our approach differed to the company in several ways. First, we used a frequentist model selection method to narrow down a wide range of polynomial terms and combinations to a smaller subset that would be taken through to Bayesian estimation. Second, as a result of that, we used a broader set of polynomial functions at model selection stage than the company used in any one meta-analysis, but used comparatively fewer at the model comparison stage. Third, we considered both fixed effects and random effects models in our analysis, but did not consider first-order polynomials. A final point of difference is that we included updated datacuts for a range of trials, including CheckMate 9ER. We also used expert

elicitation to guide in selection of curves. As a result, the EAG and the company chose different fractional polynomial distributions for each outcome, limiting direct comparability of findings.

However, a number of points merit discussion. For all-risk OS, the company's survival curves referenced against the CheckMate 9ER trial showed a similar set of curves, grouped as sunitinib and pazopanib with all other curves forming a second group. However, comparing the EAG's models to the company's models over 60 months, the EAG's fitted models appeared to show more variation between treatments early, with cabozantinib + nivolumab showing an advantage earlier in the time horizon that was not reflected in the company's analyses. Comparing HRs in all-risk PFS, both the company's and the EAG's analyses suggested that nivolumab + ipilimumab had a longer-term advantage over cabozantinib + nivolumab, though the EAG's analysis suggested that cabozantinib + nivolumab and avelumab + axitinib were closer together in effectiveness over the time horizon than the company's analysis indicated.

In the intermediate/poor risk group, the company's and EAG's analyses PFS analyses broadly aligned, though in the EAG's analysis, cabozantinib + nivolumab performs somewhat worse over the later time horizon as compared to cabozantinib. However, in the OS analyses, the EAG again found that cabozantinib + nivolumab had an early survival advantage that appeared to 'fade out'; in the company's analyses, curves for cabozantinib + nivolumab and cabozantinib are broadly coterminous over the time horizon.

Our proportional hazards analyses for OS and PFS aligned well with the results provided by the company. However, we considered ORR as the sum of complete and partial responses, whereas the company considered complete responses only. In addition, our NMA for discontinuation due to AEs used a random effects model and an informative prior distribution, whereas the company's model used fixed effects. This limited comparability. There was also a lack of comparability between NMAs on discontinuation outcomes. A possible reason for this is that we used updated datacuts for JAVELIN-RENAL-101, CLEAR and CheckMate 9ER. Unlike the company, we did not regard that meta-analysis of HRQoL was warranted given the available data.

3.7.7.3. Cochrane review

Next, we note a recently published Cochrane review that considered 1st line treatments for advanced RCC.[150] This review had a radically different scope than the current analysis as it sought to consider all published RCTs for this population, and used a frequentist analysis

paradigm. A total of 36 RCTs were included in this review, and NMAs were primarily estimated using random effects. In addition, searches were last undertaken in February 2022, which would have excluded a number of more recent datacuts we included for relevant trials. The NMA did, however, explore the impact of specific adverse events of clinical interest which could not be explored by the EAG within the timeframe for this appraisal (HFS, diarrhoea and fatigue) and was therefore considered useful for later us in the economic analyses.

3.7.7.4. Liao 2022

Finally, we compare our findings against the most recent NMA of 2nd line treatments identified. Our results are not comprehensively comparable. For example, our results may not be directly comparable for PFS as the definition used in Liao 2022,[77] 'time duration of disease progression, treatment cessation or end of the 2nd line treatment', did not align with ours, which specifically focused on time to radiological disease progression or death. Liao 2022 included nine trials, of which we regarded two as irrelevant due to not testing relevant comparators. Liao 2022 also included 2nd line data from two crossover trials, which we did not include in this NMA as second-period (i.e. post-progression comparisons) are not randomised. We also included TIVO-1 and analysed subgroup estimates where available, which this NMA did not; included BERAT and TIVO-3; and used more recent datacuts for CheckMate 025. However, where findings used similar datacuts, our NMA results were aligned.

3.8. Ongoing studies

Six relevant ongoing studies which have not year reported were identified prior to receipt of company data, including two from the trial registries search. These were:

- NCT05012371, which compares lenvatinib + everolimus against cabozantinib in a 2nd or 3rd line context after progression on a PD-1/PD-L1 checkpoint inhibitor[151];
- SUNNIFORECAST, which compares nivolumab + ipilimumab in combination against standard of care in a 1st line context in advanced non-clear cell RCC[152];
- A Study to Compare Treatments for a Type of Kidney Cancer Called TFE/Translocation Renal Cell Carcinoma (tRCC), which compares axitinib + nivolumab against nivolumab and against axitinib in a population with multiple lines[153];
- Cabozantinib or Sunitinib Malate in Treating Participants With Metastatic Variant Histology Renal Cell Carcinoma, comparing each treatment in a population with multiple lines.[154]
- REFINE, which is investigating an extended schedule for nivolumab following nivolumab + ipilimumab (8 weekly rather than 4 weekly) and is expected to produce results in 2025[155]

- A Study of Subcutaneous Nivolumab Monotherapy which is expected to complete in March 2025[156]

Three of these studies focus on the effectiveness of treatments in people with non-clear cell histologies. The NCT05012371 study is due to complete in April 2023 and is expected to provide highly relevant information on the comparative effectiveness of two treatments available for a previously treated population including data on their effectiveness after progression on a PD-1/PD-L1 checkpoint inhibitor, which is current standard practice. Unfortunately, however, this is a relatively small Phase 2 study (estimated enrolment of 90 participants). The other two studies looking at the mode and frequency of administration of nivolumab and could have a significant impact on the cost and cost-effectiveness of treatments for RCC when they report in 2025.

An additional ongoing study (RAMPART[157]) which started in 2018 was noted during clinical expert consultation as a UK study collecting information on the outcomes of adjuvant treatments in patients with a high or intermediate risk of relapse. The trial was set up to collect data on durvalumab, durvalumab + tremelimumab and active surveillance. We have been informed by clinical experts taking part in the structured expert elicitation exercise that data was also collected on patients who received adjuvant pembrolizumab in later phases of the trial. The primary completion date is noted as July 2024. It is unclear whether data is being collected within this study on the impact of treatments in the systemic, rather than adjuvant, setting and if they are when these data are likely to report.

3.9. Conclusions of the clinical effectiveness evidence

3.9.1. In relation to the decision problem and the company's submission

In the assessment of the clinical effectiveness evidence, the EAG scrutinised the company's submission, which included the CheckMate 9ER trial for first-line treatment in the target population. The EAG broadly agreed with most decisions taken by the company, but disagreed on the full range of appropriate comparators, the relevance of time to next treatment, and the importance of risk group-specific analyses. While the EAG regarded the trial as having high risk of attrition bias, the EAG also noted that the availability of 44-month follow-up was a potential strength. The EAG noted a number of potential issues with respect to generalisability of the trial (including high rates of treatment after progression) but was satisfied that the trial presented

evidence of effectiveness of cabozantinib plus nivolumab as compared to sunitinib across key outcomes, including OS, PFS and ORR. However, the EAG noted some evidence of effect modification by risk group for OS and PFS in particular, with favourable risk groups experiencing less effectiveness than intermediate and poor risk groups. Based both on the trial and on network meta-analyses (discussed below), the EAG agreed that overall, cabozantinib plus nivolumab is an effective treatment for first line RCC relative to existing treatment options and may be a consideration for patients in any risk group where a combination treatment is considered appropriate.

3.9.2. In relation to the EAG's syntheses

The EAG undertook its own SLR and identified 24 trials, of which 17 were prioritised for analysis. Collectively, the EAG's syntheses suggested that combination therapies (IO/TKI and IO/IO) were most effective at first-line, although they were also associated with high rates of adverse events, including a high rate of adverse events leading to discontinuation in the first-line setting. In the fractional polynomial NMAs, cabozantinib plus nivolumab, cabozantinib monotherapy, nivolumab plus ipilimumab, and avelumab plus axitinib all performed better than sunitinib in both the overall risk and intermediate/poor risk populations at first line. At second plus line in the overall risk population, lenvatinib plus everolimus, nivolumab monotherapy and cabozantinib monotherapy performed best. While proportional hazard analyses suggested that IO/TKI combinations outperformed IO/IO combination (nivolumab plus ipilimumab), this was not borne out in the fractional polynomial analyses.

However, despite the number of treatments available for RCC across lines and risk groups, the EAG considered that the evidence base in RCC was highly limited. With the exception of older treatments, shown in analyses to be less effective (e.g. sunitinib and sorafenib in the first line and everolimus monotherapy in the second plus line), most newer treatments were supported by only one trial. There was variation in some outcomes across trials that was not readily explained by known effect modifiers, and the EAG therefore concluded that there are some concerns about the comparability of effects across the evidence base. This is further magnified by evidence from observational sources suggesting that outcomes have improved over time, above and beyond the impact of any specific treatment. The paucity of evidence prevented statistical exploration of inconsistency in NMA and restricts confidence in any patterns in effect across potential effect modifiers. Moreover, many of the included trials conducted subgroup analyses to investigate patterns in treatment effect across risk subgroup and in the NHS,

clinicians frequently alter management according to risk category. However, analyses by risk group were limited due to the small sample sizes and a reduction in the availability of trial data (particularly in the favourable risk population). Overall, the EAG considered that there was a high degree of uncertainty in the clinical effectiveness results.

A further consideration for the clinical effectiveness results was that there was evidence of non-proportional hazards across outcomes, meaning that the results of proportional hazard NMAs are likely to be unreliable for some comparisons; at the same time, fractional polynomial NMAs were highly uncertain due to similar deficiencies in the evidence base. The narrative synthesis was also conducted based on hazard ratios that assumed proportional hazards, or on effects reported at a single follow-up timepoint, and therefore these findings may also be unreliable. Fractional polynomial NMAs were feasible for OS and PFS and suggested a different pattern of results than the other analyses. For example, while pembrolizumab and lenvatinib emerged as one of the strongest treatments across outcomes and risk groups (albeit with imprecision around the treatment effect) based on the proportional hazards analyses and the narrative synthesis, plots of hazards over time showed that this effect was being driven by a large effect in the short-term that then reduced (and even reversed) with longer follow-up; conversely, fractional polynomial NMAs produced results for pembrolizumab and lenvatinib biased in the other direction. Fractional polynomial NMAs were not conducted in the first-line favourable risk population due to data limitations.

Additional outcomes were narratively synthesised, including duration of response, time on treatment and health-related quality of life. These outcomes were not reported for all treatments and were generally restricted to analyses in an overall risk population. In the first line in an overall risk population, nivolumab plus ipilimumab, cabozantinib plus nivolumab, pembrolizumab plus lenvatinib and avelumab plus axitinib all showed a longer duration of response relative to sunitinib. The findings reported for time on treatment were not considered to be informative due to sparsity of data. No treatments were found to offer meaningful benefits for HRQoL over their comparators. In general, HRQoL was found to decrease following treatment irrespective of treatment received, and relative differences between treatments in overall response were not borne out in meaningful differences in HRQoL.

Going beyond challenges with the evidence base itself, the presented syntheses leave open a number of questions, with the most pressing relating to histology and prior treatments. First, most trials were restricted to people with clear cell RCC, which is known to have improved

treatment outcomes compared to non-clear cell histologies. The licence for cabozantinib plus nivolumab, similar to other combination treatments, does not restrict use in people with non-clear cell RCC, though the CheckMate 9ER trial was also restricted to those with clear cell disease. Based on the studies identified as part of this appraisal, there is little understanding of how treatment effects may vary in people with alternative histology RCC although the EAG does not an increase in trials being conducted in this area. Second, we were unable to explore the importance of adjuvant pembrolizumab on outcomes within this appraisal, given the availability of evidence. Clinical advice to the EAG is that receipt of adjuvant pembrolizumab may be beneficial for the population in general, but that it may reduce the benefit exhibited in subsequent treatments involving IOs. This may be particularly true in the favourable risk population, since more low risk patients can be identified in the routine scanning after adjuvant pembrolizumab.

Clinical advice to the EAG and consideration of relevant evidence highlights that optimal treatment sequencing following novel treatments at first line (i.e. IO/IO or IO/TKI combinations) remains an area of uncertainty. An exploration of the role of prior treatments in subsequent treatment outcomes will be conducted as part of Phase 2 of this appraisal, however, the evidence base appears relatively sparse.

3.9.3. In relation to real-world evidence

The EAG identified a number of real-world evidence sources and completed full assessments of quality for four sources. The EAG ultimately determined that the UK RWE dataset provided the most robust and relevant natural history data for use in an economic model. Median PFS data from the UK RWE was consistent with those reported in clinical trials, though median OS from UK patients was generally shorter than was reported in the trials. On the basis of the baseline characteristics reported or the UK RWE, the EAG was unable to identify meaningful differences in data sets that may influence OS, and this was not a primary aim within the remit of this appraisal. In general, evidence based on RCTs is considered to lack external validity due to the artificial procedures used in the trials relative to clinical practice, and a tendency for trials to exclude people with higher risk or more complex disease. The EAG considered it plausible that treatment effects, both in terms of absolute survival and relative effects, reported in the clinical trials would therefore vary from those that would be seen in clinical practice. Where appropriate and feasible, learnings from RWE will be integrated into Phase 2 of this pilot.

4. COST-EFFECTIVENESS MODEL DEVELOPMENT

4.1. Published cost-effectiveness studies

4.1.1. Search strategies

Systematic searches of the health economic literature were undertaken to identify 1) economic evaluations of relevant interventions and comparators, 2) studies reporting quality of life data in the form of utilities, and 3) UK cost and resource use studies. Search strategies are provided in Appendix A.

Search strategies were developed by an information specialist and the final strategies were peer reviewed by another information specialist within our team. The search strategies used relevant search terms, comprising a combination of indexed keywords (e.g., Medical Subject Headings, MeSH) and free-text terms appearing in the titles and/or abstracts of database records and were adapted according to the configuration of each database. No publication status (published, unpublished, in-press, and in-progress) limits were applied.

Alongside the Medline and Embase searches detailed below, the following databases were searched to identify general economic studies: INAHTA, CEA registry, ScHCARRHUD, NHS EED, EQ-5D documents, and the NICE website. All were searched from 2009 (aligning with the publication of the first NICE appraisal in RCC) to 2023. We also searched RePEc via EconPapers. Given the lack of an export functionality in EconPapers, we reviewed the first 30 hits online. Finding no unique, in-scope citations among these 30, we added no documents from RePEc.

Abstracts and titles of references retrieved by the electronic searches were screened by two reviewers for relevance against the criteria specified in Table 54. Full paper copies of potentially relevant studies were then obtained and assessed for inclusion by two reviewers using the pre-specified inclusion/exclusion criteria. At both stages, discrepancies were resolved by discussion, with involvement of a third reviewer, where necessary. All duplicate papers were double checked and excluded.

Included studies were extracted by one reviewer into a bespoke database for each search. The quality of cost-effectiveness studies evaluating cabozantinib + nivolumab was assessed using the Philips 2004 checklist for decision analytical models.[158]

Table 54: Inclusion and exclusion criteria for economic studies

PICOS item	Include	Exclude
Population	Studies of participants with advanced (stage 3 unresectable and stage 4) RCC	Studies of participants with early stage (not advanced) RCC
Intervention (economic evaluation searches only)	Cabo+nivo, pazo, tivo, suni, cabo, nivo+ipi, pem+lenv, axi, lenv+evero, evero, nivo, ave+axi*	Any other treatments not listed under inclusion Treatments used in the adjuvant setting
Comparator (economic evaluation searches only)	Any of the other interventions listed above (i.e. head-to-head studies) Usual care / physicians' choice / best supportive care	Any other treatments
Outcomes	Economic evaluations Incremental Cost Effectiveness Ratio expressed as cost per life year gained or cost per QALY Cost savings (cost-minimisation studies only) Utility studies Quality of life data expressed in the form of utilities regardless of the method of elicitation and valuation Cost and resource use studies Resource use data from UK studies Cost data from UK studies	Studies not reporting an included outcome
Study design	Economic evaluations (cost-effectiveness, cost-benefit, cost-consequence or cost-minimisation) Systematic reviews of economic evaluations or utilities Conference abstracts will be included unless data are superseded by another conference abstract or full journal article	Abstracts with insufficient methodological details Editorials and commentaries
Data limits	Economic evaluations: 2009 Utility studies: 2009 Cost and resource use studies: 2017	

Abbreviations: QALY, quality-adjusted life year; RCC, renal cell carcinoma

Notes: * as belzutifan was included within the NICE draft scope it was included within the search terms for the searches conducted, these studies will, however, not be included during screening

4.1.1.1. Searches for economic evaluations

Searches for economic evaluations were carried out in Medline and Embase, using the SIGN economics filter.[82] The same terms were used for the economic evaluation searches as for the clinical RCT searches in respect of the population and interventions. Searches were limited to 2009 onwards, aligning with the publication of the first NICE appraisal in RCC. No limits by language were used.

Conference abstracts were included for the following conferences: American Association for Cancer Research, American Society of Clinical Oncology, American Urological Association,

European Society for Medical Oncology, European Association of Urology, Genitourinary Cancers Symposium, International Conference on Translational Cancer Medicine and The International Society for Pharmacoeconomics and Outcomes Research.

4.1.1.2. Searches for health utilities

The utilities searches in Medline and Embase used the same population terms, but no intervention terms were used. Rather, population terms were combined with the CADTH utilities filter.[75] As with the economic evaluations search, searches were limited from 2009 onwards, and the same conferences were included as above. No language limits were imposed.

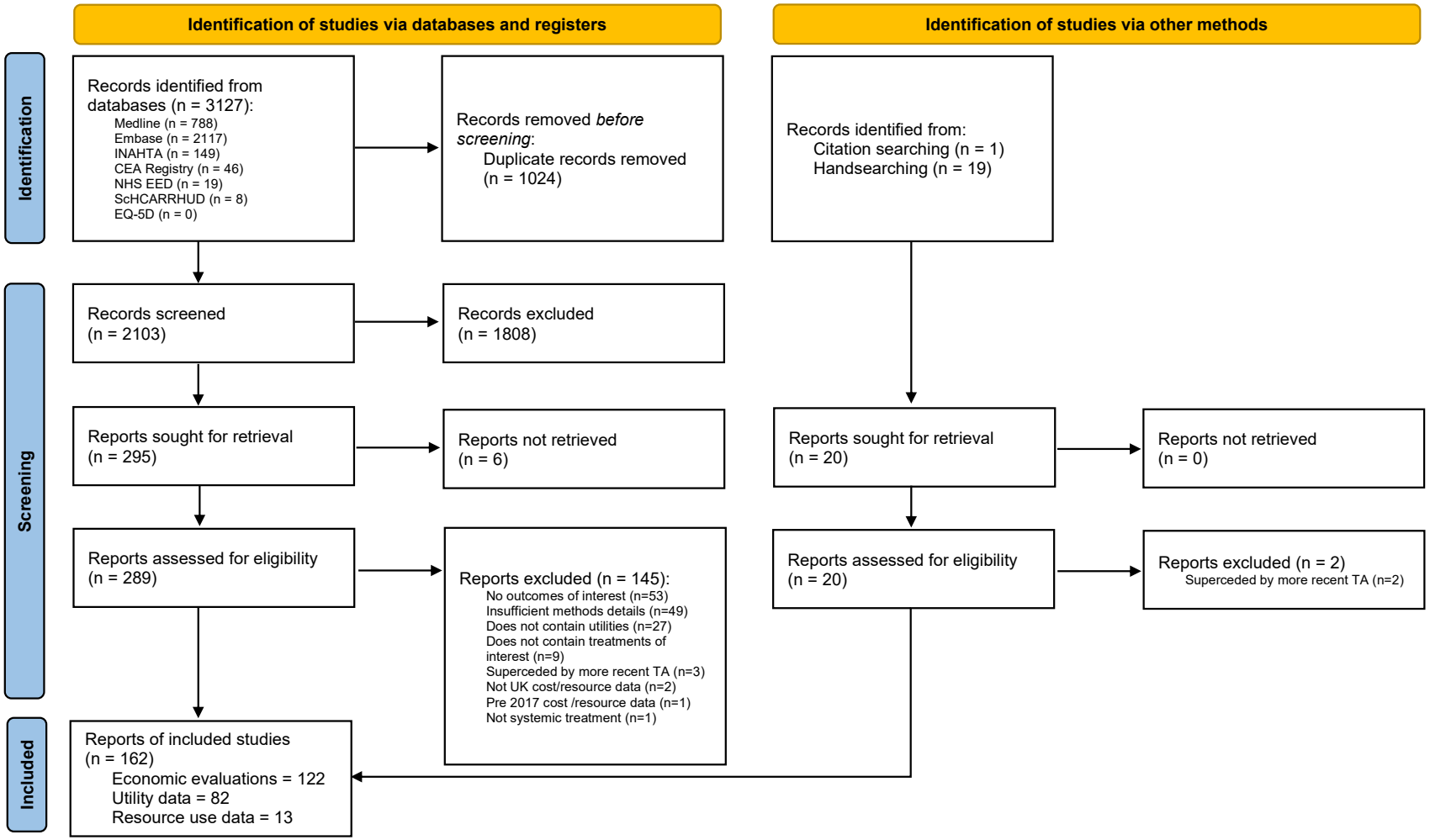
4.1.1.3. Searches for UK cost and resource use studies

UK cost and resource use searches in Medline and Embase combined population terms with the Cochrane cost of illness filter[159] and the NICE UK filter.[83] Studies were included from 2017 onwards, to ensure that only relevant data are found (aligning with the entry of immuno-oncology options into clinical practice post TA417[58]). Again, no language limits were imposed.

4.1.2. Results of the searches

In total, 162 papers were identified across the three searches (Figure 33). Some publications contained information relating to more than one review. 122 papers containing relevant economic evaluations were identified, 82 papers were identified containing utility data (discussed in Section 4.3.7) and 13 containing cost and resource use data (discussed in Section 4.3.8)

Figure 33: Economic literature review PRISMA



Abbreviations: INAHTA = International Network of Agencies for Health Technology Assessment; NHS EED = The NHS Economic Evaluation Database; ScHCARRHUD = School of Health and Related Research Health Utilities Database. Note: a number of studies qualified for more than one of the economic reviews and therefore the total across each of the 3 reviews (122 + 82 + 13) sums to more than the number of reports included (n=162)

Of the 122 economic evaluations identified, the EAG prioritised inclusion within this report to the following types of studies:

- Previous NICE technology appraisals from 2017 onward – 10 included
- Systematic reviews of cost-effectiveness studies from 2017 onward – two included
- Studies evaluating cabozantinib + nivolumab – seven included
- Sequencing models – six included
- Western (Europe, US, Canada, Australia and New Zealand) studies by recency of data – 44 included

The data extraction grid can be found in Appendix D. Data was extracted into the pre-specified data extraction sheet by one reviewer and 10% of records were checked by a second reviewer.

4.1.2.1. Learnings from previous technology appraisals

Table 55 provides a summary of economic evaluations used in previous NICE technology appraisals in RCC.

The vast majority of previous NICE technology appraisals used a simple three-state partitioned survival (PartSA) model based upon progression status. This aligns with company preferences for oncology modelling as discussed in TSD 19.[160] The use of this structure may not, however, have been ideal as within a number of these appraisals (TA780,[48] TA650,[47] TA645,[46] TA512,[51] TA417[58]) the Committee raised concerns around the way that subsequent therapy was accounted for, expressing a clear preference that costs and effectiveness of subsequent lines should match and that Committee preference was to use UK data for both. This type of analysis would be very difficult to achieve in a PartSA model without access to patient-level data for all involved treatments to allow statistical adjustment of OS. Within a state transition model, although evidence gaps would still remain, there is the flexibility to test the impact of different assumptions rather than having unquantifiable, and sometimes unacknowledged, uncertainty relating to the mismatch between subsequent treatments within trials and practice and the impact of this on effectiveness.

Another issue identified within previous appraisals relates to inconsistency in the evidence base. Different trial arms have been used to represent the reference treatment across appraisals and previous appraisals generally used HRQoL from the trial for the treatment currently being appraised. There are therefore different estimates of baseline risk for progression, death and HRQoL being used for the same population and same treatment across appraisals.

The EAG also note that the evolution of appraisals within RCC and lack of use of a common model and set of comparators has already led to some potentially counterintuitive decisions. Specifically, TA780[48] (a CDF re-review) did not compare nivolumab + ipilimumab to cabozantinib (the only other option available specifically for intermediate and poor risk disease) as it was not in scope of the original appraisal in line with standard process at the time. TA858[38] then found nivolumab + ipilimumab not to be cost-effective versus cabozantinib with pembrolizumab + lenvatinib recommended on the basis of cost-effectiveness versus nivolumab + ipilimumab and not cabozantinib due to high levels of usage of nivolumab + ipilimumab in practice.

Table 55: Summary of previous technology appraisals

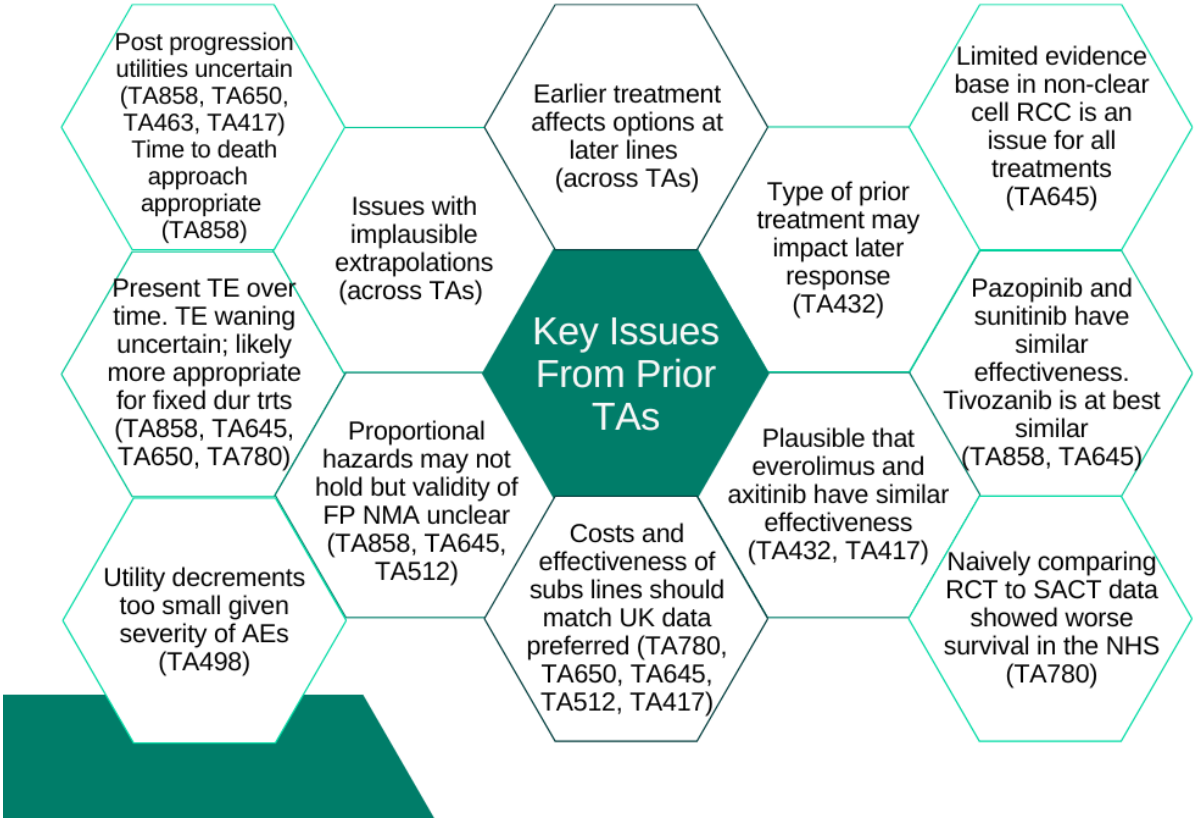
TA	Year	Recommendation population	Model type	Intervention	Comparators in final analysis	Source of HRQoL data	Committee ICER
TA858 (MTA)[38]	2023	1L int/poor risk, where nivo+ipi would otherwise be offered	3 state PartSA	Pem+lenv	Int/poor risk: cabo, nivo+ipi Favourable risk: suni, pazo, tivo	CLEAR	EAG vs nivo+ipi = £133,362 vs cabo = £166,249 (list price analyses) Not c/e vs cabo
TA830[39]	2022	Adjuvant: increased risk of recurrence after nephrectomy	State transition: DF, LR, DM and death	Pem	Routine surveillance	KEYNOTE-426 for advanced RCC	NA
TA780[48] (CDF review of TA581)	2022	1L int/poor risk	6 state PartSA (prog and tx states, terminal care, death)	Nivo+ipi	Suni, pazo	CheckMate 214	vs suni = £25,897 - £36,041 vs pazo = £24,653 - £34,132
TA650[47]	2020	1L (not recommended)	3 state PartSA	Pem+axi	Pazo, suni, tivo, cabo (int/poor risk)	KEYNOTE426	vs suni = £59,292 - £76,972 vs cabo = £29,835 - £38,346
TA645[46]	2020	1L	3 state PartSA	Ave+axi	Pazo, suni, tivo, cabo (int/poor risk)	JAVELIN Renal 101	Company: vs suni = £26,242 vs pazo = £29,542 vs tivo = £9,220 vs cabo = Dominant
TA542[52]	2018	1L int/poor risk	3 state PartSA	Cabo	Suni, pazo	TA512	vs suni = £37,793 vs pazo= £48,451 vs suni = £31,538
TA512[51]	2018	1L	3 state PartSA	Tivo	Suni, pazo	TIVO-1 trial	Pazo dominates tivo & suni
TA498[56]	2018	1 prior VEGF, ECOG 0-1	3 state PartSA	Lenv+evero	Axi, cabo, evero, nivo	AXIS	Company: vs axi = £32,906 vs cabo = £16,083

TA	Year	Recommendation population	Model type	Intervention	Comparators in final analysis	Source of HRQoL data	Committee ICER
							vs nivo = £17,146 vs evero = £96,403
TA463[55]	2017	Prior VEGF	3 state PartSA	Cabo	Axi, nivo	METEOR and AXIS	Redacted
TA432[57]	2017	Prior VEGF	State transition 4 states: stable disease (no AEs), stable disease (AEs), prog and death	Evero	BSC, axi - exploratory analysis	Swinburn et al., (2010)	vs BSC = £51,700 - £52,261 vs axi = Dominant (list price)

Abbreviations: AE, adverse event; BSC, best supportive care; CDF, Cancer Drugs Fund; DF, disease free; DM, distant metastases; EAG, external assessment group; ECOG, Eastern Cooperative Oncology Group; ICER, incremental cost effectiveness ratio; int/poor, intermediate / poor risk using IMDC criteria; LR, loco-regional recurrence; MTA, multiple technology appraisal; prog, progression; PartSA, partitioned survival analysis; tx, treatment; VEGF; vascular endothelial growth factor; vs, versus; 1L, 1st line

Notes: he ICERs provided in this table are those described within the Final Appraisal Document where possible. Where this was not possible the EAG or company ICER is provided based upon what was available and which the Final Appraisal Document appeared to most closely align to

Figure 34: Summary of issues from prior NICE appraisals of technologies for RCC



Abbreviations: AE, adverse event; FP NMA, fractional polynomial network meta-analysis; RCC, renal cell carcinoma; RCT, randomised controlled trial; SACT, systematic anti-cancer therapy dataset; TA, technology appraisal; TE, treatment effect

Figure 34 provides a summary of the key issues raised in prior NICE technology appraisals of technologies for RCC. Many of these are interlinked and stem from difficulties with the evidence base available in terms of maturity of information for extrapolation, quality of data for more historic treatments, lack of data in risk status subgroups, lack of data for non-clear cell histologies and methodological disagreements over the most appropriate way to handle violation of proportional hazards within trials.

The importance of subsequent therapy is highlighted in that earlier treatment affects options at later lines, as discussed in Section 2.4.2 and that there is some evidence that type of prior treatment may impact outcomes at later lines. It is clear from a number of prior TAs that Committee preference is for cost and effectiveness to match when considering subsequent treatments and for UK patterns of subsequent therapy to be used above trial data.

It is also clear that there are limitations to the available HRQoL data in RCC, in particular difficulties capturing the full impact of issues with tolerability for certain treatments and uncertainty around post progression utilities (a wide range of estimates are available which is likely influenced by changing practice around subsequent treatment and by collection of post progression utilities being limited to 30 days in a number of the trials).

Lastly, appraisals that have included UK RWE have shown worse outcomes in NHS practice than in trials, based on naïve comparison. There was some suggestion that this may be due to a higher proportion of patients having intermediate / poor risk status in practice than may be included in some trials.

4.1.2.2. Learnings from systematic reviews

Two systematic reviews of the cost-effectiveness of treatments for RCC were identified.[161,162] Both considered only the cost-effectiveness of immune checkpoint inhibitors.

Verma (2018)[161] identified three studies considering the cost-effectiveness of nivolumab versus everolimus for previously treated RCC[163-165]: two PartSA models with considerable differences in results (ICERs of $51,714 per QALY [pharma sponsored] and $146,532 per QALY) and driven by differences in extrapolation techniques, and a state transition model that reported a similar ICER versus everolimus to the more conservative of the PartSA approaches, but concluded that nivolumab was not cost-effective versus placebo. Uncertainties were raised in the review around optimal dosing and duration of immune checkpoint inhibitors and the impact of late presenting toxicities.

Philip (2021)[162] identified 23 studies published between 2008 and 2020, across 9 different countries (1st line treatment (n = 13), 2nd line treatment (n = 8), and 1st line and beyond (n = 2)). The majority, fourteen studies, included the use of novel immune checkpoint inhibitors nivolumab, ipilimumab, pembrolizumab), half of which found that checkpoint inhibitors were more cost-effective when compared to oral systemic therapies (sunitinib, everolimus, axitinib, pazopanib, and cabozantinib). The review did not identify any studies of cabozantinib + nivolumab and did not look in detail at the drivers of results.

4.1.2.3. Learnings from economic evaluations of cabozantinib + nivolumab

Seven publications reported an economic evaluation of cabozantinib + nivolumab (Table 56).[166-173] All of the publications used data from CheckMate 9ER (with the majority using the March

2020 database lock). The four papers not sponsored by industry compared to sunitinib. The other three compared to a variety of treatments including TKIs and combination therapies.

All five publications not sponsored by Ipsen, including the abstract sponsored by Bristol Myers Squibb (BMS), concluded that treatment was not cost-effective based upon the stated prices. BMS concluded that their wholly owned combination (nivolumab + ipilimumab) dominated when compared to cabozantinib. Conversely, Ipsen concluded in their two analyses that when comparing cabozantinib + nivolumab to nivolumab + ipilimumab, that QALY gains were either the same or the opposite direction (i.e. favouring cabozantinib + nivolumab). The rationale for these differences is unclear.

None of the publications were conducted from a UK perspective and none were high quality, with survival extrapolation methods either unclear or driven only by visual and statistical fit. Quality assessment was conducted using the Phillips checklist and is included in Appendix D.

One study explored the difference a state transition vs a PartSA model structure made upon outcomes and concluded that there was little difference. Drug costs, quality of life and effectiveness inputs were key drivers in the majority of models with relative dosing intensity (RDI) also being a key driver in one. The utility sources used by the authors of the papers that were not industry funded were acknowledged as not ideal as EQ-5D data from CheckMate 9ER was not available to them.

Table 56: Summary of published economic evaluations of cabo + nivo (1)

	Li 2021	Liao 2021	Liu 2022	Marciniak 2022
Analysis country	US	US	US	France
Funder	US government	Chinese government	Chinese government	Ipsen
Price year	2021	2021	2021	Unclear
Time horizon	Lifetime	Lifetime	10 years	50 years
Comparators	Suni	Suni	Suni	TKIs+ and combinations¥
Model structure	DES based on PFS, discontinuation & mortality due to AEs, lifetables and OS during BSC Curve selection not justified	3 state PartSA Extrapolation methods unclear	3 state models: state transition & PartSA Curve selection statistical and visual fit only	3 state PartSA Curve selection statistical fit only
Source of efficacy data	CheckMate 9ER (March 2020 DBL), AXIS, TIVO-3, dovitinib vs sora RCT[59,85,99,104]	CheckMate 9ER (March 2020 DBL)[59]	CheckMate 9ER (March 2020 DBL)[59]	CheckMate 9ER[59] (Sept 2020 DBL) NMA for comparators
Price of cabo 60mg / nivo 240mg	$491.30 / $6,849.84 (average CMS sale price)	$866.51 / $8,015.04 (Red Book)	$515 / $7,432 (average CMS sale price)	Not reported
Utilities	By line 0.82, 0.77, 0.66, and 0.494 -0.157 for Grade 3+ AEs	PFS cabo+nivo 0.848, PFS suni 0.73, progressed 0.66	PFS cabo+nivo 0.75, PFS suni 0.73, progressed 0.66	Not reported
Utility sources	Cella 2018 (METEOR)[174] De Groot 2018 (PERCEPTION)[175] Wan 2019 (CheckMate 214)[176] Patel 2021 (myeloma) [177] Wu 2018 (VEG105192 trial)[105]	Wan 2017[165] Wan 2019[176] Wu 2018[178] Data not from CheckMate 9ER. Selection methods unclear	Cabo+nivo estimated from FKSI Wan 2019[176]	CheckMate 9ER

	Li 2021	Liao 2021	Liu 2022	Marciniak 2022
	Selection methods unclear			
Subs therapy	Axi→sora→BSC	Unclear, average cost	CheckMate 9ER	Taken from individual publications for 1L therapies, includes treatments not available in the UK
Perspective	Payer	Payer	Payer	Not reported but appears to be payer
Base case ICER	$508,987/QALY	$863,720/QALY	$555,663/QALY vs $531,748/QALY*	Uses placeholder costs for some inputs 7.4 life years, 5.4 QALYs for both nivo+ipi and cabo+nivo Life-year range, 5.1–6.2; QALY range, 3.8–4.6 for TKIs Life-year range, 6.3–7.1; QALY range, 4.7–5.2 for other combinations
Key drivers	Patients age at treatment, 1L utility, cost of nivo	PF utility, cost of cabo, effectiveness parameters	PF utility, drug costs	Not reported

Notes: * state transition vs PartSA; +TKIs included: cabo, pazo, tem, tivo, sorafenib, suni; ¥ combinations: nivo+ipi, axi+ave, axi+pem, lenv+pem

Abbreviations: AE, adverse event; BRL, Brazilian Real; BSC, best supportive care; CMS, Centers for Medicare and Medicaid Services; DBL, database lock; DES, discrete event simulation; FKSI, Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index; ICER, incremental cost effectiveness ratio; OS, overall survival; PF, progression free; PFS, progression-free survival; PartSA, partitioned survival analysis; QALY, quality-adjusted life-year; RDI, relative dosing intensity; US, United States

Table 57: Summary of published economic evaluations of cabo + nivo (2)

	Tempelaar 2022	Wang 2022	Yoshida 2022
Analysis country	France	China	Brazil
Funder	BMS	Chinese government	Ipsen
Price year	2020	2022	Unclear
Time horizon	15 years	20 years	Unclear
Comparators	Nivo+ipi, pem+axi, pazo, suni	Suni	Nivo+ipi, pazo, suni
Model structure	3 state PartSA Extrapolation methods unclear	3 state PartSA Curve selection statistical and visual fit only	3 state PartSA Extrapolation methods unclear
Source of efficacy data	CheckMate 9ER Multi-dimensional treatment effect NMA vs suni	CheckMate 9ER (March 2020 DBL)	CheckMate 9ER[59] (datacut unclear) NMA for comparators
Price of cabo 60mg / nivo 240mg	Not reported	$491.20 / $3,482.57	Not reported
Utilities	Not reported	PFS cabo+nivo 0.848, PFS suni 0.73, progressed 0.66 -0.157 for Grade 3+ AEs	Not reported
Utility sources	CheckMate 9ER French value set	Li 2021, Liao 2021	CheckMate 9ER
Subs therapy	Not reported	CheckMate 9ER	Clinical studies, source and data not reported
Perspective	All payer	Health system	Not reported
Base case ICER	Cost-efficiency frontier was only comprised of two treatments: pazo and nivo+ipi. Nivo+ipi strictly dominated cabo+nivo (incremental Euros / incremental QALYs: 63,792/-0.221)	$292,945/QALY	vs suni BRL 365,591/QALY vs pazo BRL402,944/QALY vs nivo+ipi BRL347,698/QALY (int/high risk)
Key drivers	Multi-dimensional treatment effect NMAs	Drug costs, utilities at progression, subsequent treatment	RDI, discount rate, drug costs

Notes: * state transition vs PartSA; ⁺TKIs included: cabo, pazo, tem, tivo, sorafenib, suni; ¥ combinations: nivo+ipi, axi+ave, axi+pem, pem+lenv

Abbreviations: AE, adverse event; BRL, Brazilian Real; BSC, best supportive care; CMS, Centers for Medicare and Medicaid Services; DBL, database lock; DES, discrete event simulation; FKSI, Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index; ICER, incremental cost effectiveness ratio; OS, overall survival; PF, progression free; PFS, progression-free survival; PartSA, partitioned survival analysis; QALY, quality-adjusted life-year; RDI, relative dosing intensity; US, United States

4.1.2.4. Learnings from previous sequencing models

Six publications were identified that provided information on three models considering sequencing within RCC. One model looked specifically at patients assessed as IMDC intermediate / poor risk. Five of the models were discrete event simulation analyses (two papers discussed what appeared to be the same model using the DICE framework[179,180]). One model used a state transition structure.[181] Model structures varied with the more complex manufacturer led models including response, TTD, reason for discontinuation (AE or progression), TTP or next treatment, adverse events and death and the academic-led model considering only treatment line, adverse events and death.

One of the studies used data collected retrospectively from a patient registry,[182] in the Netherlands the others used trial data supplemented by network meta-analysis or trial data alone. None of the studies considered the full network included in this analysis, none report a UK perspective. Only one study considered sequencing after cabozantinib + nivolumab.[181] Key considerations within the publications include:

- **Access to patient-level data**: the majority of the models were produced with industry sponsorship and included analysis of patient-level data from manufacturer sponsored trials. This was necessary to produce the required risk equations accounting for the impact of population characteristics and prior treatments on prognosis. Where data were not available, information from treatments with a similar mechanism of action was generally substituted or additional analyses were required to calibrate the model to account for missing parameters
- **Issues with reporting of time to treatment discontinuation and time to receipt of subsequent treatments** meaning that assumptions were required (e.g. assumption of similar relative effectiveness to PFS or assumption that TTD and TTP are equal)
- **Difficulties in matching observed treatment effects for subsequent treatments** in the CheckMate 214 trial with data observed in clinical trials for subsequent therapies
- Analysis based on CheckMate 025 **assumed that the efficacy of 2nd line treatment was not affected by the 1st line agent** received (due to the 1st line options modelled being only TKI monotherapy). The model which included cabozantinib + nivolumab[181] also appeared to make this assumption although the exact source of effectiveness data was not clear
- The **2nd line treatment** preferred and most frequently observed in the trials following 1st line IO/TKI combinations other than cabozantinib + nivolumab was **cabozantinib**. After cabozantinib + nivolumab this was **lenvatinib + everolimus**
- The need to **include non-RCC mortality separately,** as trial-based mortality hazards were often decreasing at the end of trials
- The **potential for a treatment free interval** for patients receiving immuno-oncology treatments in the 1st line setting (demonstrated in a proportion of participants in CheckMate 214)

- Difficulties using standardly assessed progression to determine treatment failure on immuno-oncology due to the potential for **'pseudo-progression'**; a well-recognised phenomena that has been discussed in multiple NICE TAs, and fitting curves to PFS due to initial **drops in the KM curve linked to scanning protocols**
- Limitation in the number of lines of treatments explicitly modelled (maximum of 2 active treatments)
- **Differences between real-world treatment practice and best practice** as detailed within guidelines. In de Groot 2017[182] only 54% of the patients received a targeted therapy; one in four fulfilling eligibility criteria did not receive targeted therapy

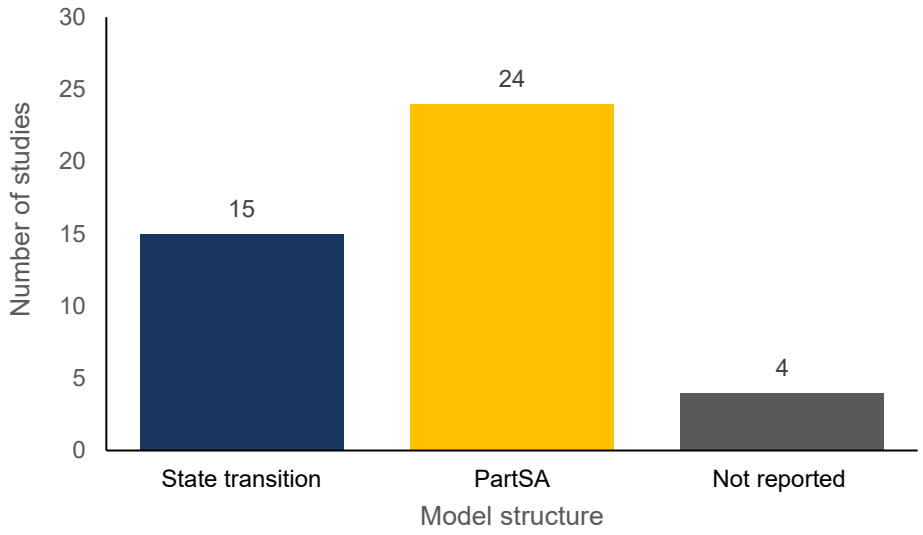
Key prognostic factors identified within a number of analyses included:

- Risk score (MSKCC)
- Age – relatively small impact
- Region (US vs Canada/West Europe/North Europe vs rest of world) – inconsistent direction of effect
- Race – inconsistent direction of effect
- Performance status (KPS, WHO, ECOG) – higher is poorer prognosis
- Histology – non-clear cell poorer prognosis
- Prior nephrectomy – improved prognosis
- Site of metastases – liver and lung metastasis poorer prognosis
- Number of lesions – more is poorer prognosis
- Laboratory values (abnormal values poorer prognosis); LDH, Alkaline phosphatase, haemoglobin, neutrophil count, albumin
- PD-L1 status (poorer prognosis for TKIs, not predictive for immuno-oncology in CheckMate 214)

4.1.2.5. Learnings from other published economic evaluations

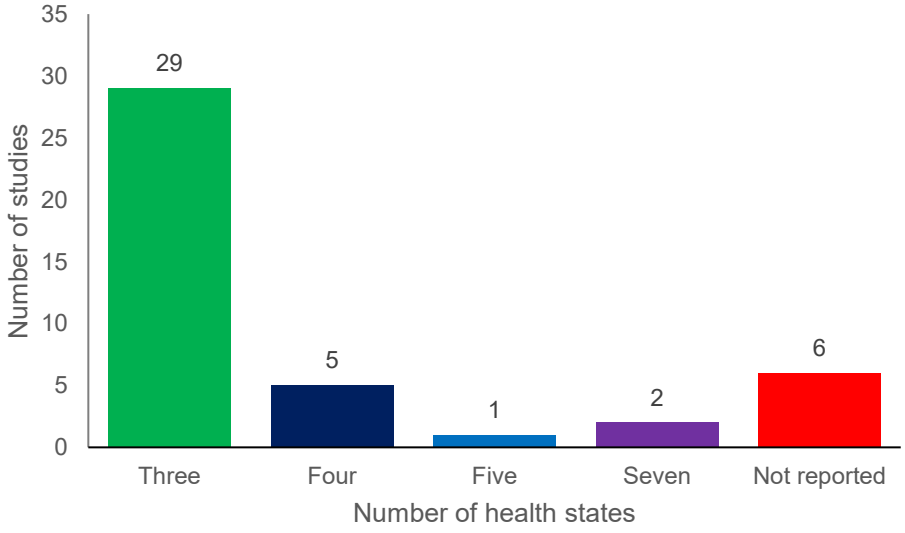
Data were extracted from 43 additional studies. 26/43 (60.5%) of the studies looked at 1st line therapies, 17/43 (39.5%) investigated 2nd line therapies. All of the studies were based in North America, Europe, Australia, or the UK. All studies either evaluated patients with poor/intermediate risk status (IMDC) or did not report the risk status. All the model structures used in these studies have been used by a previous NICE TA, literature review, or a sequencing model. The model structures used have been summarised in Figure 35 and Figure 36. All clinical effectiveness and utility inputs were derived from trials, or from previous NICE TAs.

Figure 35: Model structure used in published economic evaluations.



Abbreviations: PartSA, partitioned survival analysis

Figure 36: Number of health states used in published economic evaluations.



Models that incorporated only three states included pre-progression, post-progression and death. For those with four states, the additional health state was either progression to 2nd line treatment or progression to BSC, or they were not reported in the study. The study including five states included pre and post progression on and off treatments, and death, and the two studies with seven health states included pre-progression (no treatment), pre-progression (treatment), pre-progression (dose reduction), unobserved progression, progression detected by CT scan,

death from RCC. Both of those studies by (Raphael, 2017,2018[183,184]) seem to discuss the same state transition model evaluating the cost effectiveness of from the perspective of the Canadian healthcare system.

Sixteen 1st line studies looked at combination therapies, 14 of those studies contained nivolumab + ipilimumab which resulted in the highest QALYs gain against other comparators in all of them. The study by Zhu et al, 2023[185] evaluated two combinations: lenvatinib + pembrolizumab and lenvatinib + everolimus; both combinations resulted in similar QALY gain. Yfantopoulos et al, 2022[186] evaluated pembrolizumab + axitinib, which is outside of the scope of this appraisal, which resulted in better outcomes compared to sunitinib.

In the comparative analysis of monotherapies, cabozantinib consistently demonstrated a greater gain in quality-adjusted life years than sunitinib across all studies. Pazopanib yielded a slightly higher number of QALYs than sunitinib, albeit by a negligible margin of less than 0.1 in all studies except one which used real-world evidence (Nazha 2018[187]), where sunitinib exhibited better performance. For the 2nd line, cabozantinib came in top place in the evaluations found, followed by nivolumab, which led to a higher QALY gain than everolimus, which then had a higher QALY gain than axitinib.

There were no additional learnings relevant to the specification of the model for the pathways pilot identified in the papers reviewed.

4.2. Structured expert elicitation

4.2.1. Rationale for structured expert elicitation

The maximum follow-up available within the available clinical trials identified is just over seven years (CheckMate 025[188]). A median of 44 months of data are available for CheckMate 9ER, with the median OS only just reached for cabozantinib + nivolumab within published evidence identified so far.[120] Whilst this is relatively long when compared to the length of follow-up usually available within a NICE technology appraisal, this is nevertheless still short when compared to model time horizons of 40 years in the more recent published examples for 1st line treatments. Given this and the fact that recent changes to the treatment pathway are expected to impact on outcomes we plan to conduct a structured expert elicitation exercise to inform expected long-term survival (see Section 4.2.4).

The objective of the elicitation exercise was not to seek a 'single best answer' or point estimate from each expert, but to elicit a probability distribution representing their judgement about the relative likelihood of different values. That is, the distribution represents an expert's uncertainty based upon their existing knowledge. We sought to understand the uncertainty around the average (mean) value and not to understand individual patient heterogeneity.

Materials from the STEER repository[189,190] which was developed in line with the Medical Research Council (MRC) protocol,[189] were used to plan and conduct this exercise. The full protocol can be found in Appendix J.

4.2.2. Expert recruitment

We initially sought to recruit a minimum of five and a maximum of ten oncologists, or urologists who treat RCC, who we would expect to be the experts most likely to be able to provide input on expected survival for given treatment sequences. Following initial conversations with two urologists this criteria was narrowed to oncologists who were considered to be the speciality most able to provide information on systemic treatments.

We sought to include experts from centres from a mix of geographies across England and from a mix of types of centres: e.g. academic vs clinical, urban vs rural populations. Experts were identified by hand searching RCC publications and NHS websites. Recruitment was focussed on substantive skills as recommended within the MRC protocol rather than normative skills. We aimed to minimise conflicts of interest where possible. In particular we did not recruit experts involved in the CheckMate 9ER trial. Experts were required to declare any potential conflicts as consistent with NICE policy.

The inclusion criteria for experts were:

- Willing and able to participate within the required timeframe
- Absence of specific personal and financial conflicts of interest
- Published within the field of advanced RCC or referred by another included expert
- At least five years of experience treating people with advanced RCC

Nine experts were recruited from a total of 38 experts contacted. Expert recruitment was complicated by the junior doctors and nurses strikes which took place during the key recruitment period and the general level of business within the NHS. This led to a much higher number of

contacts being required to find experts able to participate and the timeframe for the expert elicitation exercise needing to be pushed back. In addition, during the training exercises which took place in May the clinicians requested a further delay to allow evidence from ASCO which was held 2-6 June 2023 to be provided in the background information and considered in their responses.

All nine experts completed both the training and the survey. Despite attempts to gather input from a range of geographies the majority of the experts were based in the South of England (3 in London, 2 in the South West, 2 in the East of England, 1 in the South East of England and 1 in Scotland). The mean number of years of experience treating people with advanced RCC was 15 (range 5 – 25) and the mean number of advanced RCC patients treated per year was 190 (range 20 – 600). Five of the nine experts came from a cancer research centre (Glasgow, Belfast, Cambridge, Royal Marsden, Leeds, Manchester, Oxford and Wales); all of the experts stated that their centre either had an academic focus, was a university hospital or a tertiary teaching hospital. Two of the experts stated that their population coverage included rural as well as urban geographies.

4.2.3. Quantities of interest

We sought to understand the expected PFS and OS outcomes for people receiving different subsequent therapies in UK practice, the impact of different types of 1st line treatment on PFS and OS, and the impact on OS of different sequence lengths for subsequent treatments.

There were two potential methods to elicit the required information considered, either:

- landmark survival estimates for treatment sequences; or
- landmark estimates of either PFS or TTNT per line of therapy

Based upon expert input, the latter was expected to be more intuitive and avoids issues with treatment effect being highly dependent upon subsequent therapies. Treatments to include have been selected to reflect both the CheckMate 9ER trial and UK best and current practice as described by the elicitation exercise participants.

Data were elicited for no more than 10 sequences or treatments per expert to keep the exercise manageable. Focus was given when assigning experts to each treatment to the intervention that will be first appraised using the pathways pilot model (cabozantinib + nivolumab) and they key comparators for that treatment.

Table 58: Treatments included within the expert elicitation exercise per clinician

Line	Risk Group	Treatment	Clinician number
1L	Int / poor	Cabo+nivo	1-5,6,7,8,9
1L	Int / poor	Nivo+ipi	1-5,9
1L	Int / poor	Pem+lenv	1-5,9
1L	Int / poor	Pem+axi*	2,3
1L	Int / poor	Cabo	1-5,10
1L	Int / poor	Suni	1-5,6,7,8,9
1L	Favourable	Suni	1-5,9
1L	Favourable	Cabo+nivo	1-5,9
1L	Int / poor	Pem+axi*	2,3
1L	Favourable	Ave+axi	3,6,7
2L	All risk	Cabo (after nivo+ipi)	1,6,8
2L	All risk	Cabo (after an IO / TKI combination)	1,6,8
2L	All risk	Lenv+evero (after an IO / TKI combination)	6,7,8
2L	All risk	Nivo (after 1L TKI monotherapy)	6,7,8
2L	All risk	Cabo (after 1L TKI monotherapy)	6,7,8
2L	All risk	Tivo (after nivo+ipi)	2,7,8
3L	All risk	Lenv+evero (after nivo+ipi and cabo)	4,7,8
3L	All risk	Axi (after an IO/TKI combination and cabo)	4,7,8
4L	All risk	Evero	5,7,9
Last line	All risk	BSC – from the timepoint that the patient and clinician decide that further active treatment is not desired Here we would ask about OS rather than PFS	1,4,9
Adjuvant	Int/poor	Suni	5,6,9
Adjuvant	Int/poor	Cabo+nivo	5,6,9

* This treatment is not within the scope of this pathways pilot and was included at the request of 2 of the experts involved who considered that this should be reappraised when axitinib is available in generic form which they considered would occur in the next few years. It is not in scope of the initial NICE appraisal using this information.

We had planned to provide experts with the demographics of the population to be estimated to reduce the potential for variation driven by patient characteristics. We had planned to match this to the expected UK patient population eligible for 1st line treatment, rather than to the sample in CheckMate 9ER. However, these data were not received in time. We therefore had to consider how to handle this problem:

- Ask experts to estimate for the patient population they see in practice (potential for variation but more observable for participants)
- Ask experts to estimate for the CheckMate 9ER trial population

Given the former option matches more closely to the desired decision problem population and is easier for the experts to observe and therefore comment on we asked experts to provide estimates for the population they see in practice. This was worded within the web tool as: "Please provide your estimates for the advanced RCC patient population in England (including non-clear cell where eligible for the same systemic therapies)." Information was provided on which histologies can be treated with the same treatment options as clear cell.

Experts were asked to estimate landmark PFS at three timepoints for each sequence. These timepoints were selected based upon input from Dr Larkin on the maximum amount of time patients are likely to remain progression free for most treatments and information on the available trial data for each treatment. The timepoints were presented to all of the experts during training and were considered to be reasonable:

- 3 years
- 5 years
- 10 years

For last line (BSC) we asked about OS at six months, one year and two years.

In the preliminary assessment report we specified that additional questions may be added to estimate the expected effect of adjuvant pembrolizumab per NHS guidance on OS in the advanced setting. The level of uptake of adjuvant pembrolizumab in UK practice shown in RWE will drive whether or not these questions will be required. At the time pembrolizumab was appraised uptake was expected to start at 20% of the eligible population rising to 65% in 5 years.[39] This was considered sufficient to warrant questioning.

These questions were asked as a modification of the landmark estimates for two key treatments to account for people who have received prior adjuvant pembrolizumab: nivolumab + cabozantinib (initial intervention of interest) and sunitinib (common comparator in the trials). Experts were asked to provide estimates for the intermediate/poor risk population as this represents the majority of treated patients.

As background information we provided the experts with information extracted from relevant clinical trails. We focussed the information provided on the most recent studies including the treatments considered most relevant within RCC: for 1st line patients we had initially planned to include CheckMate 9ER, CheckMate 214, CLEAR and Javelin Renal 101 and for 2nd line-plus patients CheckMate 025, METEOR and NCT01136733. Following requests from clinicians during training we added KeyNote 426, trials with non-clear cell histology (SWOG1500 and BERAT) and a short summary of the trial for adjuvant pembrolizumab (KeyNote 564). We provided the experts with:

- PFS Kaplan Meier plots for all of the treatments
- OS Kaplan Meier plots for all of the treatments

And summary tables including:

- OS and PFS HRs for all of the trials (including by risk group and 1st line)
- Baseline demographics for all of the trials
- Information on how progression was assessed within the trials

A definition of PFS was provided as follows: "the proportion of patients who have not progressed according to RECIST criteria, received a subsequent treatment or died at a particular timepoint from the start of that line of treatment. "

- Please ignore tumour flare
- We are aware that a small proportion of patients experience oligo site progression which can be treated with radiotherapy (e.g. SABR) without switching treatment. Please count these patients as progressed

This definition was included and discussed with clinical experts during the training sessions with the two clarifying bullets being added based upon recommendations provided by experts.

We included a short narrative on potential differences between assessment of progression in practice and within trials as context. Scan frequencies and definitions of progression can differ substantially between trials and clinical practice. Both Dr Challapalli and Dr Larkin also informed us that in a small number of cases patients continue being treated beyond RECIST-assessed progression on detailed review of the scan if this is considered to be of clinical benefit. This is observed in the dataset supplied by Dr Challapalli. These differences frequently lead to PFS appearing higher in real-world data than in trials whilst OS is lower in real-world data than in

trials. We therefore asked experts to assess PFS in the context of when they think the progression would occur according to RECIST rather than use a definition more aligned with practice which is impacted by less frequent scans and occasional continuation of treatment beyond RECIST-assessed progression. The wording used was: “We are aware that scan schedules and assessment criteria used for progression can differ between trials and practice. Please consider trial-like assessment (RECIST, 6-12 weekly scans) when making judgements.” 6-12 weekly was selected as broadly representative based upon the clinical evidence review (Section 3.3.2.4).

For each sequence we asked the experts:

- *1. “ What proportion of patients will be both alive and progression free at **3 years** for the advanced RCC patient population in England if they received **XXX** at **XXX** line in **XXX** risk group and had **not had previous treatment with adjuvant pembrolizumab**”*
- *2. “ **Of those patients who were alive and progression-free at 3 years**, what proportion would you expect to remain alive and progression free at **5 years** for the* advanced RCC patient population in England *if they received **XXX** at **XXX** line in **XXX** risk group and had **not had previous treatment with adjuvant pembrolizumab**”*
- *3. “ **Of those patients who were alive and progression-free at 5 years**, what proportion would you expect to remain alive and progression free at **10 years** for the* advanced RCC patient population in England *if they received **XXX** at **XXX** line in **XXX** risk group and had **not had previous treatment with adjuvant pembrolizumab**”*

The second and third questions are formatted in such a way as to make them conditional on the answer to the first question in order to account for dependence between the parameters.

For the questions related to use of adjuvant pembrolizumab we asked:

- *4. “*Your previous answer for patients receiving **XXX** at 1st line in the intermediate / poor risk group estimated the number of people who would be alive and progression-free at **3 years** when they had NOT received adjuvant pembrolizumab. How many do you think would be alive *if they **HAD received adjuvant pembrolizumab more than 12 months ago**?”*
- *5.* “Your previous answer for patients receiving **XXX** at 1st line in the intermediate / poor risk group estimated the number of people who would be alive and progression-free at **5 years**, **of those who were alive and progression free at 3 years**, when they had NOT received adjuvant pembrolizumab. How many do you think would be alive if they **HAD received adjuvant pembrolizumab more than 12 months ago**?”
- *6.* “Your previous answer for patients receiving **XXX** at 1st line in the intermediate / poor risk group estimated the number of people who would be alive and progression-free at **10 years**, **of those who were alive and progression free at 5 years**, when they had NOT received adjuvant pembrolizumab. How many do you think would be alive if they **HAD received adjuvant pembrolizumab more than 12 months ago**?”

These questions were piloted with Dr Larkin who suggested three changes:

- Amend the wording around patient population from “English patients” to “patient population in England”
- Remove the qualitative question: “Would you expect the impact of adjuvant pembrolizumab on OS and PFS to be similar across risk groups? Please detail why / why not and if not what you expect the difference would be?” as risk group isn’t assessed until relapse and most relapsers will be by definition favourable risk as they are likely to be picked up earlier due to frequent scanning associated with adjuvant treatment
- Focus questions on adjuvant treatment to the favourable risk group for the reason suggested above

The first two of these suggestions were implemented when sending out the surveys. The final suggestion was not implemented due to space limitations and updates to higher-priority clinical issues in other domains.

For all of the estimates provided we asked the experts to provide the rationale for their answers and any comments.

4.2.4. Approach to elicitation

Given the timeframe available, the following approach was used to seek quantitative expert input:

- One-to-one or group meeting to introduce the exercise and provide training; the training was adapted from the PowerPoint slides provided within the STEER tools and included background materials for each of the trials (see Appendix J)
- Online survey to sent to experts 19.06.2023 for remote individual completion within 2 weeks using the roulette method of the STEER R tool (example https://nice-rcc-clinician-survey.shinyapps.io/rcc_r_code_clinician_1/, dummy unique identifier 0000). The tool includes:
  - Elicitation of plausible upper and lower limits (95% CI) as an initial step
  - Elicitation of values using the roulette method
  - Feedback of values for expert revision and request for provision of rationale and comment
- Check responses and follow-up queries sent if any responses are unclear or inconsistent
- Distributions to be fitted to individual expert elicited judgements – beta distribution given the information provided was expressed as proportions
- Mathematical aggregation via linear opinion pooling

There is a lack of empirical evidence on whether fixed interval methods (such as the roulette method) or variable interval methods work better for healthcare decision making, and both methods have been used in this context.[189] Fixed interval methods are generally preferred by experts and are more intuitive, but there may be a tendency for experts to focus on the shape of the histogram rather than the probabilities they are expressing. Given the timeframe of this work and the number of quantities of interest necessitates conducting the elicitation via remote survey the roulette method was preferred as the benefit of increased intuitiveness was considered to outweigh the potential issues with focus.

There is also a lack of empirical evidence to inform a preferred method to fit distributions,[189] therefore we used the beta distribution which is commonly used where information is in the format of proportions and fitted distributions to each experts responses individually prior to pooling.

The MRC protocol advises the use of linear pooling with equal weights for mathematical aggregation for simplicity and due to a lack of research on how to generate appropriate weights.[189]

4.2.5. Results

All nine recruited oncologists completed the survey. Of the maximum of 270 question responses 256 (95%) were received. Three additional responses were discounted from the analysis as the clinician indicated that they had not understood the question. Three of the clinicians who completed the survey provided probabilities rather than conditional probabilities for the 5- and 10-year timepoints which required data to be reformatted prior to analysis to ensure consistency of results. The results of the exercise were then discussed briefing with Dr Larkin with his commentary provided below.

Clinician estimates from the expert elicitation exercise for sunitinib lay above the CheckMate 9ER KM curves. Contrary to trial data, our clinicians expected a higher proportion of patients to be both alive and progression free at 3 years. Cabozantinib + nivolumab outcomes were expected to be more similar to the trial. The cabozantinib + nivolumab treatment combination is not available for untreated advanced RCC patients in the UK, hence clinicians may have relied more heavily on trial data to make their progression/survival estimates in the elicitation survey. Unlike other therapies, all four clinicians that provided commentary for cabozantinib + nivolumab stated that they relied on trial data alone to make their estimation. The sunitinib estimates being

above the CheckMate 9ER trial data was unexpected. This may be in part due to the CheckMate 9ER Kaplan Meier data being at the lower end of the trial PFS KMs (results were more similar to those reported in CheckMate 9ER) and also in part due to the expectations of the clinicians included in the exercise. It was considered unlikely to be due to the clinicians coming from more academic centres as the majority of aRCC patients are treated in large academic centres. Estimates provided for other combinations lay relatively close to the trial data from the individual trials.

For all treatments where data was available in the UK RWE clinician estimates were above the observed information. Consultation with Dr Larkin suggested that one potential factor behind this could be for the combination therapies in particular clinicians may consider that they can get more out of these treatments now that there is more experience using them in an aRCC setting. In addition clinicians were asked to estimate PFS in a “trial-like” manner.

Interestingly, the type of prior treatment appeared to influence outcomes estimates. For patients receiving cabozantinib 2nd line, there was a lower proportion of patients expected to be alive and progression free at 3 years after receiving prior TKI monotherapy therapy (mean 14%; 95% CI 8% - 23%) than after nivolumab plus ipilimumab therapy (mean 29%; 95% CI 18% - 40%), or IO/TKI combination treatment (mean 31%; 95% CI 22% - 41%). One of the clinicians completing the survey noted that they would expect cabozantinib to perform less well after TKI monotherapy. Two clinicians noted they would expect cabozantinib to behave similarly following IO/IO and IO/TKI combinations. Dr Larkin noted that the activity of cabozantinib would be expected to be lower after receiving treatment with a prior TKI (particularly sunitinib, pazopanib or tivozanib) due to similarities in the mechanism of action and that this would be expected to be particularly evident following TKI monotherapy. This is not something that has been accounted for within the state transition model for this appraisal and may bias results in favour of TKI monotherapy.

The IMDC risk group influenced the outcome estimates of different types of therapies differently. For patients receiving sunitinib 1st line, clinicians estimated that 15% more patients would be alive or progression free at 3 years in the favourable risk group (31%) compared to the intermediate/poor risk group (16%). In contrast, outcome estimates for cabozantinib + nivolumab were broadly similar for patients with favourable risk (36%), and those in the intermediate/poor risk group (33%). Similarly, for pembrolizumab + axitinib the outcome estimates were similar in both favourable (34%) and intermediate/poor risk groups (27%). This

indicates that clinicians did not consider the effect size of IO / TKI combinations to be as large in the favourable risk group as for intermediate/poor risk patients. Dr Larkin considered this to be in line with expectations as patients do similarly well on ICIs regardless of risk group whereas IMDC risk groups are defined in order to be prognostic for TKIs.

There was a difference in clinician responses for patients receiving sunitinib and cabozantinib + nivolumab with or without prior adjuvant therapy. The outcome estimates for patients receiving sunitinib with prior adjuvant therapy (46%) indicated that 30% more patients were expected to be alive and progression free at 3 years compared to patients receiving sunitinib at 1st line without a prior line of adjuvant treatment (16%). Whereas 10% fewer patients were expected to be alive and progression free at 3 years when receiving cabozantinib + nivolumab with prior adjuvant therapy (23%) compared to cabozantinib + nivolumab alone without a prior line of adjuvant treatment (33%). The responses comparing outcomes with and without prior adjuvant therapy were provided by 3 clinicians who had answered both questions. One clinician made an error when completing the survey question for cabozantinib + nivolumab (with prior adjuvant therapy), so their response was excluded from the mean value in this group. Unfortunately, in the comments provided by the clinicians there was no clear rationale for the difference in expected outcomes between patients who receive a prior line of adjuvant therapy and those who do not. Dr Larkin considered the result to be in line with his expectations as for the sunitinib comparison patients will be picked up earlier if they have had a prior adjuvant therapy as they will be scanned more regularly and therefore metastatic spread will be diagnosed at an earlier and more treatable stage; whereas he would expect patients to derive less benefit from a subsequent ICI as by definition patients have demonstrated resistance to pembrolizumab even if there was a gap of at least 12 months between treatments.

Of all the 1st line therapies, the outcome estimates for nivolumab + ipilimumab demonstrated the greatest conditional survival, 67% at 5 years and 59% at 10 years respectively. Clinicians stated that they based their judgement on existing data that indicates that a relatively high proportion of these patients will be long term responders, and the expectation that patients on CTLA4 inhibitors such as ipilimumab will demonstrate a “tail of the curve effect”. Dr Larkin considered this to be in line with his expectation and did not expect a similar effect for IO/TKI combinations.

Table 59: Expert elicitation results

L	Treatment	P	Y	n	Mean	Median (95% CI)	Variance	Commentary
1	Cabo	Int/poor	3	4	17%	16% (11%, 24%)	0.001	Based on existing data (1)
1	Cabo	Int/poor	5	4	32%	32% (22%, 44%)	0.001	Looks to have a low rate of longer-term disease control for patients (1)
1	Cabo	Int/poor	10	4	18%	18% (11%, 28%)	0.001	Expect this group to be small due to progressive downward slope (1)
1	Cabo+nivo	Int/poor	3	9	33%	33% (22%, 43%)	0.004	Based on existing data (4)
1	Cabo+nivo	Int/poor	5	9	51%	50% (34%, 66%)	0.004	Based on existing data (2) There will be a gradual reduction in patients responding but at 3 years many will still be responding at 5 years (2) PFS curve plateaus between 36 and 50months so possibly few events by 60 months
1	Cabo+nivo	Int/poor	10	9	33%	28% (16%, 43%)	0.003	No plateau expected for IO / TKI (1) Uncertain (1) Large range of answers- quite difficult to predict that far ahead but there will definitely be long term responders (1) Based on existing data (3) Estimating that 1/3 to 1/2 would not have progressed between yrs 5 and 10 (1)
1	Nivo+ipi	Int/poor	3	6	32%	32% (23%, 42%)	0.003	Based on existing data (2)
1	Nivo+ipi	Int/poor	5	6	67%	67% (54%, 79%)	0.003	Based on existing data (2) High durability of responses (1)
1	Nivo+ipi	Int/poor	10	6	59%	59% (47%, 70%)	0.003	Based on expected proportion long-term responders (1) Most patients will remain in remission but there will be competing causes for mortality (1) Tail of the curve effect of CTLA4 (1)
1	Pem+axi	Int/poor	3	2	27%	26% (22%, 31%)	0.001	Based on existing data (1)
1	Pem+axi	Int/poor	5	2	60%	60% (48%, 73%)	0.000	Based on existing data (1)
1	Pem+axi	Int/poor	10	2	31%	31% (25%, 39%)	0.000	A few people with favourable disease will get longer-term control (1)
1	Pem+lenv	Int/poor	3	6	33%	33% (23%, 43%)	0.003	Based on existing data (3) Expected outcome worse than clinical trial population (1)
1	Pem+lenv	Int/poor	5	6	49%	49% (34%, 65%)	0.004	Based on existing data (2)
1	Pem+lenv	Int/poor	10	6	30%	30% (21%, 39%)	0.003	Data so far has an almost linear downward trend (1) Do not expect a high rate of longer-term responders (1) Competing causes of mortality (1) There is an expectation of maintenance of PFS with the use of PD1 inhibitors albeit not on the same magnitude as with CTLA 4i (1)

L	Treatment	P	Y	n	Mean	Median (95% CI)	Variance	Commentary
1	Suni	Int/poor	3	8	16%	14% (9%, 21%)	0.002	The KM curve looks poor at 2 years (1) Proportions of longer term non- progressors will be lower in everyday practice than in the trial (1) Unlikely that patients will remain progression free beyond 18 months, as most progress within 1 year (1) Based on existing data (2)
1	Suni	Int/poor	5	8	32%	32% (21%, 43%)	0.001	Would expect a low range to remain progression free which is difficult to predict (2) These will be mostly good prognosis patients that have done w ell with initial therapy (1) Based on existing data (2)
1	Suni	Int/poor	10	8	34%	34% (18%, 48%)	0.001	There are very few patients with longer term disease control in this group (1) It is very likely that after 10 years 2-7% remain progression free (2) Proportion of patients progressing between 60 and 120 months would be slightly higher than those progressing between 36 and 60 months (1) Based on existing data (1)
1	Ave+axi	Fav	3	3	46%	46% (38%, 51%)	0.002	Expect the favourable risk group patients to do very well (1) Based on existing data (1)
1	Ave+axi	Fav	5	3	51%	50% (41%, 60%)	0.002	Expect to see more durable responses in a proportion of patients (1) Based on existing data (1)
1	Ave+axi	Fav	10	3	32%	32% (24%, 43%)	0.002	Small proportion will achieve complete response (1) In favourable risk sunitinib is as efficacious as IO-IO or IO-TKI combination. The rate of progression beyond 60 months would also be expected to be similar between Avelumab-Masitinib and Sunitinib (1)
1	Cabo+nivo	Fav	3	5	36%	36% (29%, 43%)	0.002	Based on existing data (2) Based on trial data, however, would expect to be lower in real life. Would have expected favourable risk group to have done better in the trial (1) There is PFS benefit but not OS benefit with combination of I/O -TKI in favourable risk RCC (1)
1	Cabo+nivo	Fav	5	5	34%	37% (27%, 48%)	0.002	Due to the progressive downward slope without a plateau, I expect this to be further reduced by 30-40% (1) I think that most people will have progressed at this point (1) There is PFS benefit but not OS benefit with combination of I/O -TKI in favourable risk RCC (1)
1	Cabo+nivo	Fav	10	6	38%	51% (28%, 64%)	0.002	Considerable uncertainty but I expect progressive deterioration at 10 years (1)

L	Treatment	P	Y	n	Mean	Median (95% CI)	Variance	Commentary
								Most people will have progressed at this point and will not be o n the therapy (1) Better patients selected out at 5 years, there is reasonable possibility of another 5 years survival (1) There is an expectation of maintenance of PFS with the use of PD1 inhibitors albeit not on the same magnitude as with CTLA 4 (1)
1	Pem+axi	Fav	3	2	34%	34% (29%, 39%)	0.001	Based on existing data (1)
1	Pem+axi	Fav	5	2	56%	55% (44%, 68%)	0.001	Based on existing data (1)
1	Pem+axi	Fav	10	2	28%	28% (23%, 33%)	0.000	There are few patients with longer- term disease control in this group (1)
1	Suni	Fav	3	5	31%	31% (23%, 40%)	0.002	Based on existing data (1) Median PFS of favourable risk RCC patients on sunitinib is 4-5 years (1)
1	Suni	Fav	5	5	45%	45% (32%, 59%)	0.003	Most patient in the favourable risk group won't get longer term disease control (1) The median PFS of favourable risk RCC pts on sunitinib is 5 ye ars (1)
1	Suni	Fav	10	5	27%	27% (18%, 38%)	0.002	With a continued downward slope it is reasonable to assume approx. 40-50% of those progression free at 3 years remain progression free at 5 years (1) There are few patients with longer term disease control in this group (1) This represents the favourable risk group with a very good pro gnosis (1)
1	Cabo+nivo (with prior adjuvant)	All	3	2	23%	23% (14%, 33%)	0.003	Based on existing data (2)
1	Cabo+nivo (with prior adjuvant)	All	5	2	33%	33% (18%, 54%)	0.002	Based on existing data (1)
1	Cabo+nivo (with prior adjuvant)	All	10	2	29%	29% (15%, 51%)	0.000	Based on existing data (1)
1	Suni (with prior adjuvant)	All	3	2	46%	45% (34%, 57%)	0.003	Speculating based on the response to VEGF TKI in TKI naïve patients (1)
1	Suni (with prior adjuvant)	All	5	3	50%	50% (40%, 61%)	0.002	Would expect majority of patients on sunitinib to progress within 18 months (1) Speculating based on the response to VEGF TKI in TKI naïve patients (1)

L	Treatment	P	Y	n	Mean	Median (95% CI)	Variance	Commentary
1	Suni (with prior adjuvant)	All	10	3	33%	33% (26%, 41%)	0.001	Would expect majority of patients on sunitinib to progress within 18 months (1) Speculation based on extrapolation from patients who receive sunitinib without prior adjuvant pembrolizumab (1)
2	Cabo (after IO/TKI)	All	3	3	31%	31% (22%, 41%)	0.002	Even more difficult to predict most will have progressed at 10 years I believe this to be very similar as the situation after IO +IO (1) I would anticipate similar trends with 2L cabozantinib after IO/I O or IO/TKI combination 1L (1)
2	Cabo (after IO/TKI)	All	5	3	29%	28% (14%, 46%)	0.001	patients are less likely to remain progression free at 5 years aft er 2L therapy (1)
2	Cabo (after IO/TKI)	All	10	3	31%	31% (14%, 44%)	0.000	Very similar to cabozantinib after IO+IO Hard to predict but likely to be a small proportion with possibly a broad range (1) patients are less likely to remain progression free at 5 years aft er 2L therapy (1)
2	Cabo (after nivo+ipi)	All	3	3	29%	29% (18%, 40%)	0.004	Poor prognostic group 3 years after starting 2L I do not expect many to remain progression free (1) Not likely that there will be many patients who are progression free on 2L therapy (1) Based on existing data (1)
2	Cabo (after nivo+ipi)	All	5	3	40%	39% (21%, 59%)	0.004	Hard to predict but in the few who had been progression free some long term responders may be hiding (1) Not likely that there will be many patients who are progression free on 2L therapy (1) Based on existing data (1)
2	Cabo (after nivo+ipi)	All	10	3	36%	35% (14%, 51%)	0.001	Very rare to be progression free 10 years after 2L therapy (1) Based on existing data (1)
2	Cabo (after TKI mono)	All	3	3	14%	14% (8%, 23%)	0.002	
2	Cabo (after TKI mono)	All	5	3	34%	33% (18%, 56%)	0.002	Expect proportion will be less than cabozantinib after 1L IO combinations (1) About 20% are progression- free by 20months then the rate of events seems to plateau and would expect 1 in 10 to 1 in 6 patients not to progress (1)
2	Cabo (after TKI mono)	All	10	3	62%	63% (30%, 87%)	0.001	Less durable responses in the long term (1) Approx 15%, bell curve, slight bias to lower end (1)
2	Lenv+evero (after IO/TKI)	All	3	3	21%	20% (13%, 29%)	0.002	Likely to be a low proportion at 5 years as they have already shown TKI resistance by progressing on cabozantinib (1) Would not expect more than 10% to be progression free by 3yr s with a 3L therapy (1)
2	Lenv+evero (after IO/TKI)	All	5	3	27%	23% (10%, 42%)	0.002	Likely to be a low proportion at 5 years as they have already shown TKI resistance by progressing on cabozantinib (1)

L	Treatment	P	Y	n	Mean	Median (95% CI)	Variance	Commentary
								Would estimate at least 50% will progress from yrs 3 to yrs 5 o n a 3L regimen (1) Very low numbers now (1)
2	Lenv+evero (after IO/TKI)	All	10	3	44%	54% (15%, 81%)	0.000	Would expect greater percentage of people to progress from 5 to 10yrs than from 3 to 5yrs on a 3L treatment (1)
2	Nivo (after TKI ono)	All	3	3	25%	25% (17%, 34%)	0.003	Expect durable responses with 2L Nivo (1) Based on existing data (2)
2	Nivo (after TKI mono)	All	5	3	36%	35% (19%, 52%)	0.002	Expect durable responses with 2L Nivo (1) Based on existing data (1) Unlikely >10%, bias toward lower end (1)
2	Nivo (after TKI mono)	All	10	3	37%	37% (13%, 56%)	0.000	Expect patients to survive longer with immunotherapy (1) Likely <10% (under 5 really), skew to lower values (1)
2	Tivo (after nivo+ipi)	All	3	3	14%	9% (6%, 14%)	0.001	Less durable responses in the long term (1) Based on existing data (1) Bias to low values (1)
2	Tivo (after nivo+ipi)	All	5	3	59%	58% (34%, 74%)	0.001	People post progression on IO therapy will do poorly with later lines of therapy (1) Based on existing data (1) <10% (1)
2	Tivo (after nivo+ipi)	All	10	3	54%	56% (33%, 57%)	0.000	There will be few longer-term survivors in this group (1) The percentage not progressing will be comparable to sunitinib and cabozantinib (1)
3	Axi	All	3	2	10%	50% (1%, 52%)	0.001	Expect low proportions as axitinib is less effective than lenv+evero (1)
3	Axi	All	5	3	41%	73% (28%, 73%)	0.000	Would estimate that more than 2/3 will progress from 3 to 5yrs on a 3L therapy (1)
3	Axi	All	10	3	48%	79% (31%, 81%)	0.000	Selecting out a small number of patients (1)
3	Lenv+evero (after nivo+ipi and cabo)	All	3	3	5%	4% (2%, 8%)	0.000	Len-evero has the best PFS among available treatments, so expect a higher proportion of patients to remain progression free (1) Based on existing data (1)
3	Lenv+evero (after nivo+ipi and cabo)	All	5	3	68%	73% (35%, 60%)	0.001	Len-evero has the best PFS among available treatments, so expect a higher proportion of patients to remain progression free (1) Very low, expect <10%, bias toward <5% (1)
3	Lenv+evero (after nivo+ipi and cabo)	All	10	3	65%	65% (29%, 71%)	0.001	Very low, likely <5% (1)
4	BSC	All	0.5	2	13%	53% (2%, 58%)	0.002	Patients usually die fairly quickly particularly if they have been very TKI dependent (1)

L	Treatment	P	Y	n	Mean	Median (95% CI)	Variance	Commentary
4	BSC	All	1	3	14%	40% (6%, 46%)	0.002	Of those who are alive at 6 months a slightly better biology can be expected but it will be short-lived. Great uncertainty therefore broad range. (1) Those with more indolent disease will have survived to 6 mont hs so around 1/5 may get to a year (1)
4	BSC	All	2	3	6%	35% (1%, 37%)	0.001	Similar argumentation as before. It will be a poor prognostic group (1) This will be the indolent patients (1) Based on real world data (1)
4	Evero	All	3	3	6%	6% (3%, 12%)	0.001	Would not expect more than 1 in 10 patients to be progression free by 3yrs using a 4L treatment (1) No data to support use of everolimus in the current era of imm une check point inhibitors (1)
4	Evero	All	5	3	20%	20% (15%, 26%)	0.001	The percentage progression free at 5yrs with a 4L therapy would be lower than with a 3L therapy (1) No data to support use of everolimus in the current era of imm une check point inhibitors (1)
4	Evero	All	10	3	10%	10% (5%, 17%)	0.001	Even less patients would be progression free by 10yrs using a 4L treatment (1)

Abbreviations: IO, immune-oncology; TKI, tyrokinase inhibitor

Notes: 5 and 10 year data are conditional on survival to the prior timepoint

4.3. EAG economic analysis

4.3.1. Model structure

A *de novo* decision model was constructed for this appraisal. Adaptation of previous models, including the model used within the TA858 MTA, was not possible as these were not accessible for such use and also due to differences in the scope of this and previous appraisals.

The following factors were considered when determining the model structure to be used:

- The nature of the disease
- The need to be able to look at multiple decision nodes within the treatment pathway
- The key issues identified within the review of previous economic analysis and NICE technology appraisals
- Methodological guidance
- The available data (type, format and coverage)
- Timelines

4.3.1.1. Nature of the disease

The goal of treatment for RCC is to extend life and delay progression; with long-term survival considered a reasonable goal in the context of many active agents.[191,192]

People may go through rdimultiple lines of treatment. Experts consulted in the scoping meeting for this appraisal recommended that a maximum of four lines of treatment followed by BSC should be incorporated in the model. A previous UK audit found that on progression 69% of patients were able to receive 2nd line therapy, 34% were able to receive 3rd line therapy, 6% were able to receive 4th line therapy and only 1% received a 5th.[193]

Improving HRQoL by relieving symptoms and tumour burden is also an important clinical outcome for people with RCC.[191] Quality of life is impacted by both the stage of the disease and treatment received. Experts consulted indicated that TKI toxicities can have considerable impact on quality of life, particularly as people cannot take prolonged treatment breaks. Within the scoping workshop for this appraisal, experts noted these include chronic fatigue, chronic diarrhoea and hand / foot syndrome. With immuno-oncology treatments, immune-related adverse events are rare but can be serious in nature.

In addition to the impact on the patient, HRQoL is predictive of mortality in RCC; particularly non-RCC-specific mortality,[194] along with other well recognised factors such as age and sex.

Treatment durations vary. Treatment is either given until progression or unacceptable toxicity, or for some immuno-oncology treatments, stopping rules are in place such that treatment is only given for a fixed length of time (typically two years).

4.3.1.2. Surrogacy between PFS, TTD and TTNT

A targeted review was conducted to investigate the plausibility of surrogacy between different endpoints in advanced RCC (see Appendix F for details). The papers identified indicated that:

- RECIST-defined overall response rate and progression-free survival are not reliable surrogate end points for median OS or the treatment effect on OS in trials of PD-(L)1 inhibitor therapy[195-199]
- For targeted agents PFS is a more reliable surrogate for OS; particularly in trials which did not allow cross-over after disease progression and studies published before 2005[200,201]
- PFS may be predictive of PPS for targeted treatments at 1st line (a longer PFS meaning a longer PPS[202]); PPS is then more predictive than PFS of OS[203]
- TTNT may be a more valuable surrogate endpoint for previously untreated patients receiving PD-(L)1 inhibitor therapy[204]
- In a real-world setting prior to the wide-spread availability of IO/TKI combinations (n=171) there was a moderate correlation between PFS, TTNT and TTP with OS. The correlation coefficient for PFS and TTNT was similar (Spearman's correlation coefficients of 0.70 and 0.68)[205]. TTD, was however, less well correlated with OS (Spearman's correlation coefficient of 0.56).

Analysis from the UK real-world evidence dataset indicated a high level of correlation between TTD and PFS endpoints (Spearman's correlation coefficient of 0.83 for TTD vs PFS and 0.91 for PFS vs TTNT). Clinical expert advice to the EAG was that TTNT and PFS are well correlated and similarly TTD and PFS are well correlated for TKIs and that TTNT is a reasonable proxy for PFS. Figure 37 demonstrates that in general TTD and TTNT follow the same shape as PFS with a short lag between treatment discontinuing, progression and starting the next line of treatment (around 1 month between each).

Figure 37: PFS vs TTNT in the UK RWE

Data supplied by BMS in response to the preliminary assessment report indicate that a similar shape can be observed for both PFS and TTD for patients treated with sunitinib as rates decrease at a similar rate over time. In contrast with patients treated with nivolumab + ipilimumab, there is an increasing difference between PFS and TTD over time as a plateau appears to be forming from approximately two years for nivolumab + ipilimumab in terms of PFS whilst TTD continues to decrease.

Kaplan Meier data were requested from Ipsen in the same format for CheckMate 9ER, however, the data supplied had implemented an unexpected censoring rule (the company censored treatment with nivolumab when treatment stopped with cabozantinib and vice versa) and these data cannot therefore be used to investigate the relationship between PFS and TTD for different treatment types. The data we do have which includes TTD for both parts of the combination does not indicate the same sort of relationship (Figure 39).

Figure 38: KM curve of PFS IRRC-assessed, primary definition and TTD by treatment arm: CheckMate 214 intermediate-/poor-risk patients (60-month data cut)

Abbreviations: KM, Kaplan-Meier; PFS, progression free survival; TTD, time to discontinuation

Figure 39: KM curve of PFS versus TTD: CheckMate 9ER all risk population (44-month datacut)

Abbreviations: PFS, progression free survival; TTD, time to discontinuation

4.3.1.3. Conceptualisation of disease model

Given the above, if this model is conceptualised entirely using a disease-oriented approach, as recommended by TSD 13,[160] it would consist of health states based upon:

- Length of life
- Disease status; whether or not the patient has progressed on their current line of treatment and what line of treatment they are receiving (which may be a reasonable proxy for progression)
- Type of treatment received and whether the patient is on or off treatment
- Patient characteristics which are likely to impact upon length, and quality of life, such as age, sex and risk status should also be considered as necessary. In the case of a cohort model, it is necessary to ensure that the patient cohort modelled is reflective of UK practice and that changes in quality of life and mortality risk attributable to the aging process rather than the disease are captured.

4.3.1.4. Available data

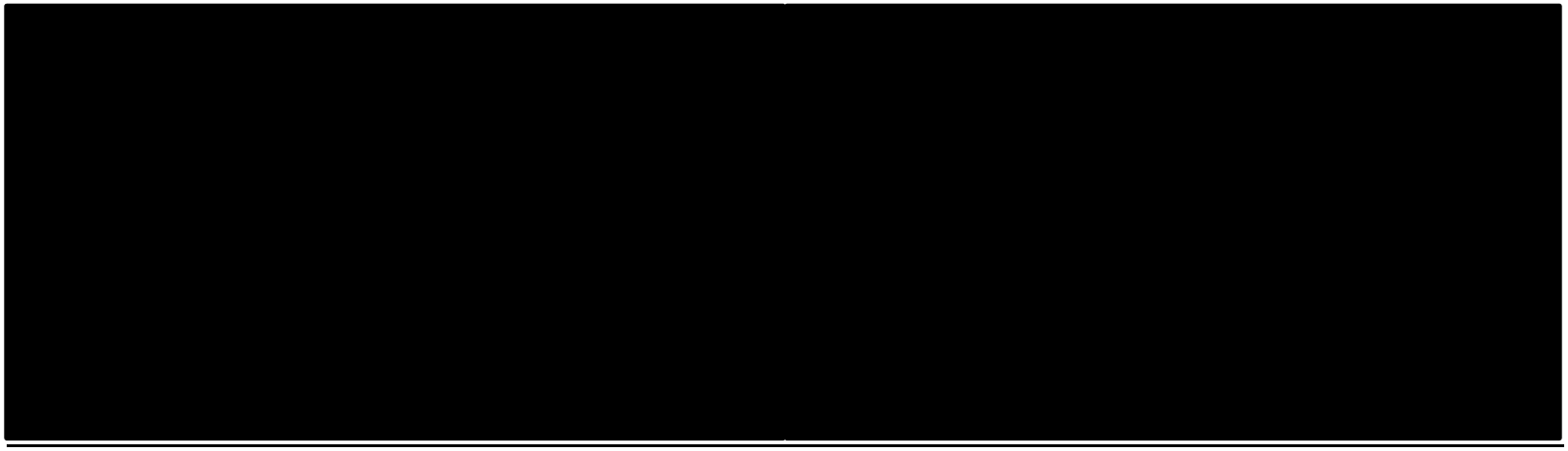
As discussed in Section 3, all identified RCTs provided information on OS and PFS endpoints and 14 of 24 trials reported HRQoL data. Only two trials reported data on TTP and relatively few reported TTD. Data for risk subgroups are less complete than for the overall population, with gaps more of an issue in the favourable risk population. Anonymised IPD was provided to the EAG for CheckMate 9ER for all endpoints except TTD by therapy type. Anonymised IPD was also provided to the EAG for 15 UK centres including OS, PFS, time on treatment (1st line only), line of treatment, risk status and other population characteristics. Data from previous modelling exercises conducted within prior NICE appraisals is not available to the EAG for model input.

It should be noted that PFS as measured within trials and PFS as measured in practice can differ substantially as patients are not routinely scanned as frequently in practice as in trials.[206,207] This can lead to PFS in the real-world appearing longer relative to OS than in trials.

Figure 40 demonstrates that when comparing the sunitinib arm in the UK RWE to that in the CheckMate 9ER trial the PFS outcomes for favourable risk patients are extremely similar whereas OS in the UK RWE is lower than in the trial. For intermediate/poor risk patients after the initial 3 months the curves separate with trial patients having more favourable PFS and for OS the difference is even more pronounced. The difference in OS outcomes between the trial and the UK RWE is expected given the strict inclusion criteria applied to trials and difference in availability of subsequent therapies across markets.

Figure 40: Comparison of UK RWE to CheckMate 9ER for suni

PFS OS



4.3.1.5. Key issues identified within previous economic analysis

The developed model should be able to handle the following additional issues identified in prior economic analyses (Section 4.1.2):

- Matching costs and effectiveness for subsequent lines of treatment
- The potential for treatment effect waning
- Lack of clarity over the most appropriate approach to modelling quality of life (progression status vs time to death).

The first of these is the most relevant to determining the overarching model structure as, although the precedent for prior appraisals has been the use of a partitioned survival approach in most previous TAs, this structure cannot readily handle adjustment for a different subsequent therapy case mix where patient-level data cannot be accessed to implement statistical analyses to adjust for treatment switching.

The latter of these is not possible for us to address as data was not provided by Ipsen for quality of life by time to death and data from prior appraisals is redacted.

4.3.1.6. The need to be able to look at multiple decision points

In order to fulfil all of the objectives, the model needs to be able to start at a user-defined line of treatment for a user-defined population and include a user-defined list of therapies available at each line from then onwards. The type of treatment received in a prior line impacts on options available at later lines and may also impact outcomes.

This sort of problem naturally lends itself to a discrete event simulation (DES) model or a state transition structure. The sequencing models identified within the economic literature review were all discrete event simulation analyses.

TSD15 considers the key benefits of a patient-level simulation to be:

- The ability to model non-linearity with respect to heterogeneous patient characteristics
- The ability to determine patient flow by the time since the last event or history of previous events
- Avoiding limitations associated with using a discrete time interval
- Flexibility for future analyses, particularly when compared to models implemented in Excel
- The ability to model interactions - not relevant to this decision problem

- Potential for efficiency savings within probabilistic sensitivity analysis (PSA)

As anonymised patient-level data in a format where patient characteristics and outcomes are able to be linked by a unique identifier are not available to the EAG for any of the treatments involved in this decision problem, the ability to model non-linearity with respect to heterogeneous patient characteristics is of no additional benefit.

A DES would be more efficient for handling time-to-event outcomes for subsequent lines of treatment where an exponential curve fit is inappropriate, however, alternatives such as the use of tunnel states are available in a state transition structure. The limitations associated with a discrete time interval can be reduced through the use of a smaller time interval.

There are also disadvantages: there can be difficulties in interpretability due to the complex nature of such models and DES models are indeed an investment; they take additional time to build compared to simpler model structures. The timeframes available for this pilot do not lend themselves to the use of a DES. For example, the IVI-NSCLC simulation model took a year and a half to build.[3]

There are a limited number of examples of use of DES within prior oncology NICE technology appraisals[208-210] and only one the authors are aware of where the disease area endpoints were OS and PFS.[208] The drivers for this are likely a mixture of precedent, data availability to gain the benefits from additional flexibilities and issues with interpretability and level of complexity for reviewers.

For example, in the abiraterone appraisal (TA387), the company submitted a DES in order to allow more flexibility to reflect a sequence of treatments and to allow the modelling of response to treatments that depend on previous treatments, both highly relevant to this decision problem. The submitted model also benefited from the availability of patient-level data allowing the modellers to account for patient characteristics that may impact on outcomes. The Committee, however, agreed that using a DES model was not unreasonable, but considered that the company's model was particularly complex.[211] The ERG considered that "an individual patient simulation by means of a DES could have been avoided, since acknowledging patient heterogeneity does not necessarily require patient-level simulation."[212]

4.3.1.7. Methodological guidance

The most relevant TSDs to consider in determining the most suitable model structure(s) for this decision problem are TSD13, TSD15 and TSD19.[160,213,214] The application of TSD13 is discussed in Section 4.3.1.1 and the application of TSD15 is discussed in 4.3.1.6. Given the majority of prior appraisals used a partitioned survival approach and those that did not use this structure were state transition models, the recommendations provided in TSD19 were given careful consideration.

TSD19 recommends consideration is given to both theoretical and practical considerations in determining modelling approach. In this case assuming that PFS and OS are independent of each other, as is the case for a PartSA analysis, would be a considerable stretch to credibility given the nature of the disease and clinical advice received. Given the data identified so far for OS (Section 3), a substantial proportion of the modelled time horizon will use extrapolated data, median OS was only just reached for CheckMate 9ER within the most recently published data-cut for example.[215] As noted in TSD19: "the lack of structural link between endpoints in PartSA models may increase the potential for inappropriate extrapolation."

There are also limitations to the implementation of a state transition structure given the limited data available in the context of this appraisal which need acknowledging. As patient-level data are not available to the EAG, a multi-state modelling approach such as that defined by Williams et al. cannot be implemented.[216] Limited data are available to define the split between progression and death events within PFS and what data are available does not provide information on the timing of events. Only two trials identified within the literature review reported data on TTP. This means that NMA is only possible for PFS as a whole at a given line of treatment rather than for individual transitions.

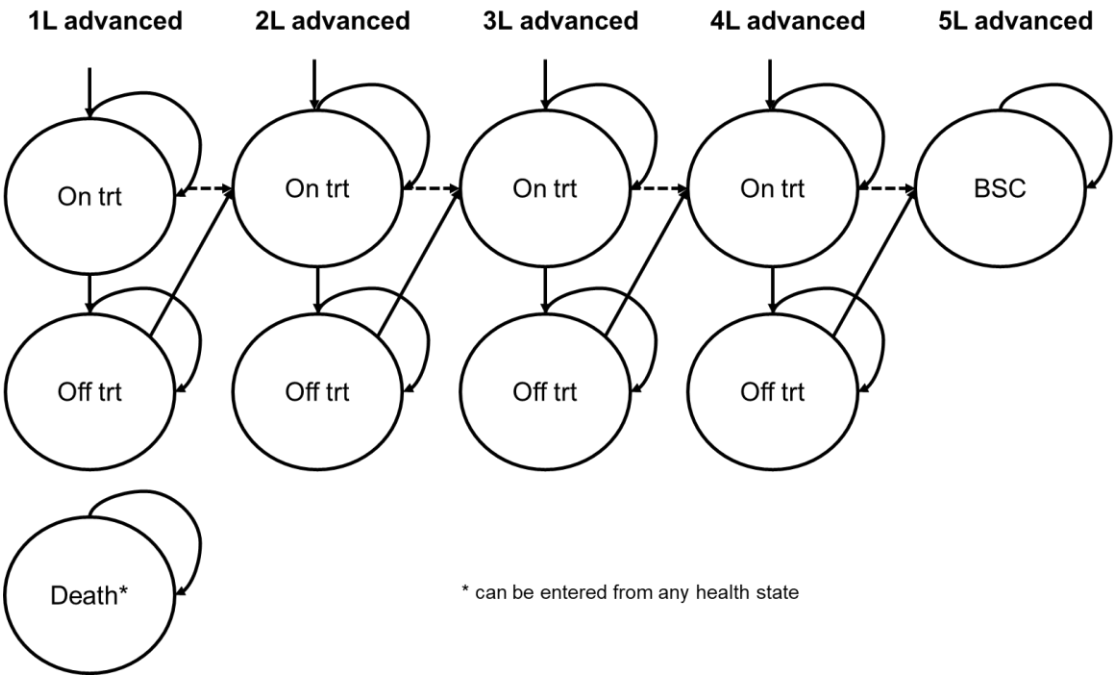
TSD19 recommends the presentation of results based upon a PartSA approach alongside those from a state transition model where a state transition structure is used given the need for further methods research.

4.3.1.8. EAG model structure

Figure 41 demonstrates the planned EAG model structure. The model is expected to allow for up to four active lines of treatment with patients who complete four lines moving to BSC. Patients will be able to receive BSC as a line of treatment at earlier lines, in this case patients will remain on BSC within that line until death.

Transitions between lines are driven by progression status. Transitions between the on and off treatment states are driven by TTD. The option to allow the use of TTNT was originally considered to make best use of data from RWE, however, in eventuality this was not required as the RWE information supplied to the EAG contained PFS.

Figure 41: EAG model structure



Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; 4L, 4th line; 5L, 5th line; BSC, best supportive care; trt, treatment

Given the various considerations detailed above, the base case model structure is a hybrid of a partitioned survival and state transition approach based upon the approach used within TA798.[217] TTP and PFS data from the UK RWE (base case) and CheckMate 9ER (scenario analysis) were extrapolated and the difference between the two used to define pre-progression survival (Pre-PS). Treatment effects for other treatments were applied from the NMA and assume that the treatment effect across TTP and PFS is the same. We refer to this hybrid simply as a state transition model throughout the rest of the report.

Data for time on treatment / time to treatment discontinuation (TTD) were also taken from the UK RWE (base case) and CheckMate 9ER (scenario analysis) and extrapolated. PFS data were used for the relative treatment effect for comparators here as well, given the lack of reported TTD data. Available data from trials which report TTD were used to check that the relationship between TTD and PFS is similar to that within CheckMate 9ER in other trials where treatments

are given until progression or unacceptable toxicity. This was the case for all treatments except nivolumab + ipilimumab where a different relationship was apparent (see Section 4.3.1.2). For fixed duration treatments, the treatment duration was capped to the maximum treatment duration in the SmPC (base case) or included in the model using the mean number of doses received based upon the relevant trial where available (scenario analysis). Relative dosing intensity was taken into account in the base case.

Effectiveness data for subsequent lines following progression on 1st line treatment were taken from available RWE for the majority of treatments with trial data used to model relative effects based upon the NMA. The proportion of patients receiving each type of treatment was modelled to reflect UK practice within the base case analysis. Tunnel states are used to track the time since entry into state for patients receiving 2nd and later lines of treatment.

The structural assumptions made within the base case model are therefore:

- OS is dependent upon progression status and line of treatment; this implies surrogacy between PFS and OS, an assumption which appears to be supported by available literature
- OS is independent of whether or not a patient is on treatment within a particular line
- TTD and PFS are independent; the impact of this is expected to be limited and will be mitigated through selection of the same functional form for fitted curves
- TTP and PFS are independent; the impact of this is expected to be limited and will be mitigated through selection of the same functional form for fitted curves
- The treatment effect from the NMAs for PFS can be applied to TTP, Pre-PS and TTD endpoints
- Patients receive subsequent treatment on progression – this is in line with how PartSA models are implemented and was considered an acceptable simplification as UK RWE showed only a relatively small difference in timing between PFS and TTNT (mean [redacted] days at 1st line)
- Transitions for 1st line are dependent upon risk status, transitions for later line patients are not dependent upon risk status (given that in practice this is only measured at 1st line)

The impact of the type of previous treatment on outcomes at later lines was included where possible, however, the ability to do this is limited based upon data identified. In particular:

- The evidence available looking specifically at the impact of sequencing of different treatments is limited
- There is no trial evidence specific to 3rd or 4th line and the 4th line data available from the UK RWE has a low sample size

- No evidence was available within the UK RWE for sequences following either nivolumab + cabozantinib or pembrolizumab + lenvatinib.

A PartSA is also presented as recommended within TSD19. This model assumes by its nature that OS, PFS and TTD are independent and that any differences between the subsequent therapy mix in practice, CheckMate 9ER and other trials within the NMA do not impact either on relative effectiveness modelled.

Given the proposed primary model structure (state transition), calibration to expected OS estimates was considered as an option. In the end this was not considered necessary as the PartSA analyses were available to cross-check against. This may be further explored in Phase 2.

4.3.1.9. Model implementation

The model was implemented in R given the complexity of the future need to evaluate large numbers of treatment sequences, the need for the model to be reusable for future HTAs and the number of structural options required to be explored.

The use of R has a number of benefits including the integration of the conduct of the core statistical analysis (survival curve extrapolation) within the model.[218,219] Table 60 provides a comparison of the analytical capabilities of R and Excel from a published example using a side-by-side PartSA and state transition structure. The advantages to run time and analytical options are clearly demonstrate for the simpler decision problem addressed by that model (only one line of treatment).

Table 60: Comparative analytical capabilities between R and Excel models in oncology

Functionality	R model	Excel model
Live fitting of parametric models	All parametric models are fitted to the active dataset	Parametric models need to be fitted to the active dataset externally, and results copied into model—a laborious task for updates to data-cut or subgroup exploration
PartSA and StateTM modelling	Model includes PartSA and StateTM modelling strategies. These are informed by the internally calculated parametric fits	Model includes PartSA and StateTM modelling strategies. These are informed by models fit outside of Excel with estimates pasted in

Functionality	R model	Excel model
PSA—time taken for 1000 PSA runs using base-case settings	1.42 min	13.2 min
One-way sensitivity analysis—time taken to run 109 parameter scenarios	0.27 min	2.4 min
Automatic report generation	Report template is set up within R Markdown to automatically populate tables and figures with active modelling analyses when selected	Highly challenging to include; not included
Quality control	Table included with selected diagnostic checks Linear code with vectors and data frames produced by single calculations that need to be checked once. However, tracing an individual calculation from start to finish can take longer than in Excel Packages used are open-source: version to be used needs to be defined to ensure stability over time	Diagnostic checks included in the patient flow sheet Cell-by-cell checks were required across all sheets because of individual calculations, meaning there was potential for drag down error and inconsistency within columns and data frames
Model size	5.1 MB—includes R scripts and Excel input workbooks containing simulated IPD, general population survival statistics and cost inputs	30.9 MB—single workbook
Version control	Managed by the version control software Git to allow tracked changes, code reversions and parallel work streams	Manual change log. Multiple versions required to allow reversions. Difficult to work in parallel

Adapted from Hart et al. R and Shiny for Cost-Effectiveness Analyses: Why and When? A Hypothetical Case Stud[219]y

Abbreviations**:** MB megabytes, MCM mixture-cure modelling, PartSA partitioned survival analysis, IPD individual patient-level data, PSA probabilistic sensitivity analysis, StateTM state transition model

The EAG, however, note that R is less familiar than Excel to many stakeholders within the NICE process. To mitigate the potential impacts of lack of familiarity on model transparency the model input sheet has been designed in Excel and intermediate outputs (patient flow) are provided in Excel. In addition NICE have commissioned the DSU to provide an independent external validation of the model code.

The model is intended to be made open-access using 'GitHub' to improve replicability and collaboration. The model was built broadly aligning with good practice guidelines, for example, the Zorginstituut Nederland National Health Care Institute (ZIN) guidelines for building models in R.[220] Underlying data (model inputs) do not need to be publicly available and can be shared confidentially with NICE abiding to the principles for handling confidential information outlined in the 2022 manual.[74]The publicly available version of the decision model which will be published following conclusion of the nivolumab + cabozantinib appraisal will use dummy data in the correct format as inputs where data are marked as either academic or commercial in confidence within the original data source. The dummy data will be created using the methods used to redact an Excel model as part of a NICE submission.

Data which are expected to need to be marked as confidential and redacted to reduce the potential for back-calculation of confidential prices include:

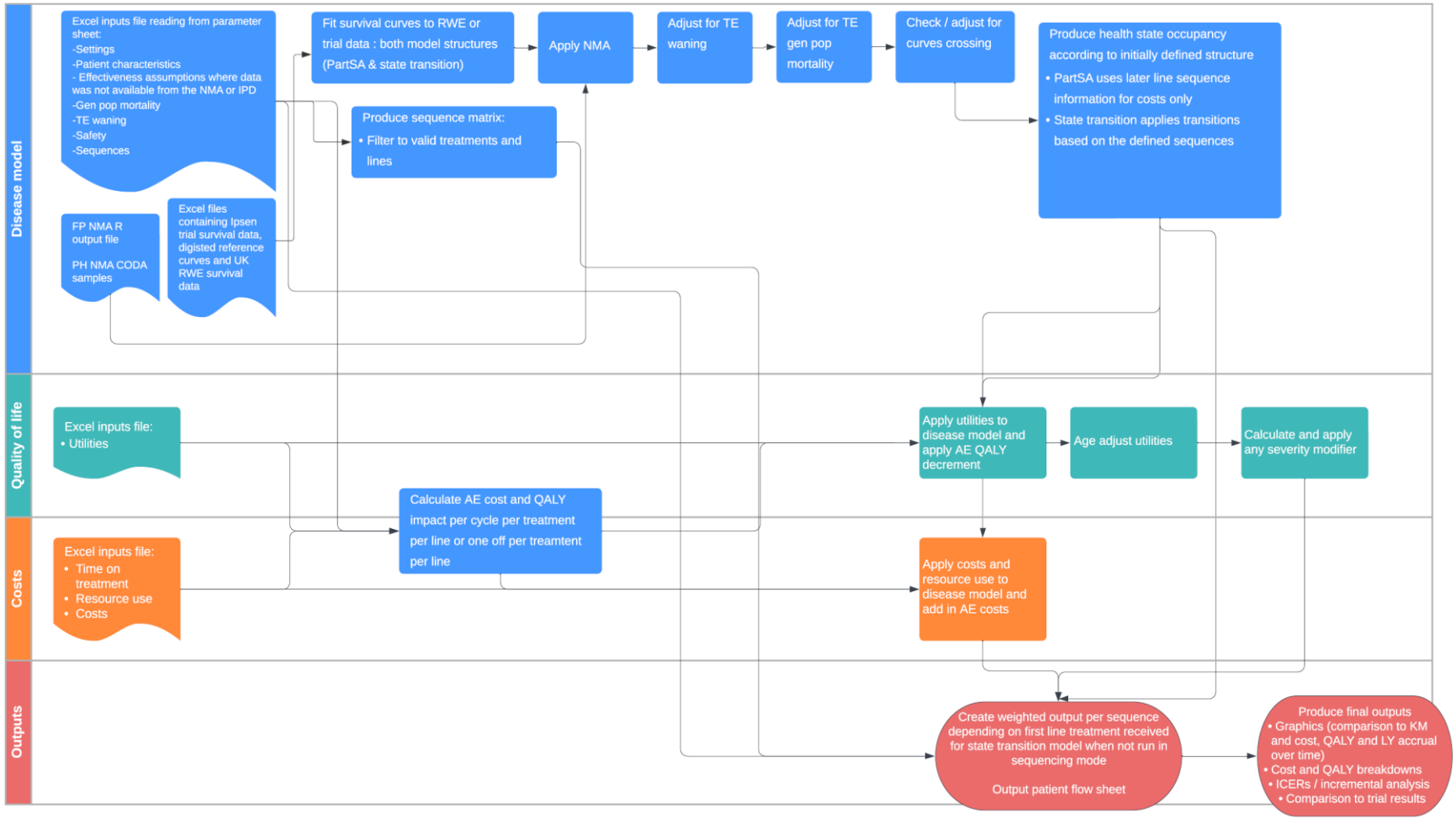
- PAS price discounts
- Any individual patient-level data provided by the company
- Time on treatment input data
- Relative dose intensity input data
- Market share data for subsequent therapies
- Reported ICERs (PAS price and list price).

A later stage of this pilot following the evaluation of cabozantinib + nivolumab will involve the incorporation of a Shiny front-end to the R model. Shiny is an open source R package enables the user to build web applications using R.[221] This will allow model users to interact via an easy-to-understand user-interface operating via their web browser.

Figure 42 demonstrates the model flow for each of the modules incorporated within the R model. Inputs to the decision model come from five sources:

- The main Excel inputs workbook which contains data and settings for the disease model, utilities and resource use and costs
- The R output file from the fractional polynomial NMA
- An Excel output file containing the CODA samples from the proportional hazards NMA
- An Excel file containing pseudo patient-level data for the reference curves for each population, treatment, trial, line and endpoint for the base case and scenario analyses; or
- The RDS output from the survival analysis (available to stakeholders for whom patient-level data access is restricted due to confidentiality)

Figure 42: EAG model flow diagram



Abbreviations: AE, adverse event; NMA, network meta-analysis; RWE, real world evidence; TE, treatment effect

The methods for each of the models required to produce the desired outputs are described in detail in the sections below.

The cost effectiveness of the interventions was estimated in terms of an incremental cost per additional QALY gained, as well as the incremental cost per life year gained (LYG), net monetary benefit and net health benefit. Base case analyses are probabilistic as this generates expected outcomes and costs and is in line with the NICE manual.[74]

Intermediate outputs including the patient flow sheet and graphical outputs such as fits to KM curves are presented, as well as the final model outputs describing cost-effectiveness and its drivers.

4.3.2. Population

The model population aligns with the decision problem population with results for the appraisal of cabozantinib + nivolumab presented for relevant treatments for untreated advanced or metastatic RCC followed by a subsequent therapy mix reflective of actual or expected UK practice.

Subgroup analysis has been presented for intermediate-/poor-risk and favourable-risk subgroups as defined in the IMDC criteria. The NICE scope requests the presentation of subgroup analysis by prior treatment. Very few patients in CheckMate 9ER received adjuvant treatment. This is not in line with the expectations for uptake of adjuvant pembrolizumab from TA830 which estimates that at full uptake 18% of patients receiving systemic therapy will have had a prior line of adjuvant treatment (see footnote of Table 61 for how this was calculated). Section 4.3.5.8 provides details of exploratory scenario analysis conducted to explore the impact of this mismatch between the available clinical trial data and expected practice.

Population characteristics were taken from the UK RWE data in the base case and CheckMate 9ER in scenario analysis (Table 61). Patients in the UK RWE were on average older than those in the CheckMate 9ER trial, other patient characteristics were similar.

Table 61: Patient characteristics included in the economic analysis

	UK RWE	**CheckMate 9ER**
% IMDC int/poor risk	77.6%	77.3%
Age: mean (SE) All risk	64.4 (0.28)	60.9 (0.41)

	UK RWE	CheckMate 9ER
Int/poor	64.2 (0.33)	61.49 (0.66)
Favourable risk	65.4 (0.56)	61.51 (0.90)
% female		
All risk	29.0%	26.1%
Int/poor	29.5%	25.5%
Favourable risk	26.5%	28.1%
Weight kg (SE)		
All risk	83.38	80.59 (0.76)
Int/poor	81.26	78.55 (0.86)
Favourable risk	90.98	87.94 (1.72)
Prior adjuvant treatment	Scenarios tested: 0%, 5.5%, 18%	

Note: scenarios for % receiving prior adjuvant treatment were calculated as the upper and lower bound of the market shares from TA830 (20 and 65%) based on the proportion of patients eligible in the UK population: 83% clear cell * 55% prior nephrectomy * 60% high risk

4.3.3. Treatments included

The treatments included within the decision model for the 1st line setting align with those specified in the decision problem (Table 3 and Figure 6).

Table 62: Treatments included within the decision model

Treatments	1L population			Administration type and frequency	Treatment duration
	All risk	Fav risk	Poor / int risk		
Cabo+nivo[62]	x	x	x	Cabo: 40mg orally once daily Nivo: 240mg every 2 weeks or 480mg every 4 weeks IV	Until disease progression or unacceptable toxicity Max 24 months for nivo
Pazo[222]	x	x	x	800mg orally once daily	Until disease progression or unacceptable toxicity[50]
Tivo[223]	x	x	x	1340 mcg orally once daily for 21 days, followed by a 7-day rest period	Until loss of clinical benefit or unacceptable toxicity[37]
Suni[224]	x	x	x	50mg orally once daily, for 4 consecutive weeks, followed by a 2-week rest period	Until disease progression or unacceptable toxicity[49]
Cabo[62]			x	60mg orally once daily	Until disease progression or unacceptable toxicity
Nivo+ipi[225]			x	Nivo: 3 mg/kg IV every 3 weeks for the first 4 doses Ipi: 1 mg/kg IV every 3 weeks for the first 4 doses	Maximum 4 cycles of combination treatment

Treatments	1L population			Administration type and frequency	Treatment duration
	All risk	Fav risk	Poor / int risk		
				Nivo maintenance: 240mg every 2 weeks or 480mg every 4 weeks IV starting 3 or 6 weeks after the last dose of combination treatment respectively	Monotherapy until loss of clinical benefit or unacceptable toxicity[37]
Pem+lenv[226,227]			x	Pem: 200mg every 3 weeks of 400mg every 6 weeks IV Lenv: 20mg orally once daily	Until disease progression or unacceptable toxicity Max 35 3 weekly cycles for pem[37] or equivalent number of 6-weekly cycles

Abbreviations: Cabo, cabozantinib; IV, intravenous; ipi, ipilimumab; lenv, lenvatinib; nivo, nivolumab; pem, pembrolizumab

For subsequent lines of treatment (which may be comprised of either active drug treatment or BSC) the EAG considered the following sources of data to determine what was included within the decision model:

- UK RWE – preferred source
- Trial data from CheckMate 9ER
- Clinical expert input to determine which sequences of treatment are valid for use in practice

Subsequent surgeries and radiotherapy were not considered as a line of treatment and were included only as a cost according to the proportion of patients expected to receive such treatment at each line.

4.3.4. Perspective, time horizon, cycle length, discounting and price year

The model uses an NHS and Personal Social Services perspective in line with the NICE reference case.[74]

The time horizon for the economic analysis was selected to be long enough to reflect any differences in costs or outcomes between the technologies under comparison. This is 40 years in line with the other recent appraisals for untreated advanced RCC TA858, TA780, TA650 and TA645.

A weekly cycle length was applied to account for the difference in dosing regimens across treatments. This is consistent with TA858, TA780, TA650 and TA645. Half cycle correction was not applied given the short cycle length.

Costs and outcomes were discounted at 3.5% per annum after the first year in accordance with the NICE manual.[74] All costs were expressed in UK pounds sterling for the 2022 price year (as the latest NHSCII inflation index was available only until 2022 during the time this report was prepared).

4.3.5. Treatment effectiveness and extrapolation

Modelling of treatment effectiveness requires extrapolation of 4 different curves for the reference treatment at each line in the model base case:

- PFS – progression and death are classed as events
  - Within CheckMate 9ER [redacted] of patients in the nivolumab + cabozantanib arm and [redacted] in the sunitinib arm were censored due to receipt of subsequent treatment (FDA censoring rules), TA858 demonstrated that use of EMA versus FDA censoring rules made little difference in another trial (CLEAR), therefore, given the low proportion and lack of impact in prior appraisals whilst this does not align with the model structure additional analyses were not requested
- TTP – progression is classed as an event and death is classed as a censor variable
- TTD – treatment discontinuation and death are classed as events
- Post progression survival (or post last line survival) for the last line of treatment – time measured starts from progression on the prior line and death is classed as an event.

Within the scenario analysis using PartSA OS, PFS and TTD required extrapolation for the reference curve at the 1st line of treatment only.

The reference treatment extrapolated for the 1st line was sunitinib given this is the comparator in the majority of the available RCTs, a treatment used in UK practice for all risk groups and the most frequently used treatment at 1st line in the UK RWE (n=326). The reference treatment for 2nd and 3rd line when using the UK RWE was as cabozantinib as this treatment was frequently used at both lines (n=245 and n= 103) and the data were mature compared to other treatments. When using trial data the reference treatment for 2nd line-plus was everolimus as this represented the treatment for which the most mature trial data was available (from CheckMate 025).

In line with the NICE manual[74] and discussion from other recent appraisals[228] data for the reference treatment was taken from UK RWE in the base case:

“Quantifying the baseline risk of health outcomes and how the condition would naturally progress with the comparator(s) can be a useful step when estimating absolute health outcomes in the economic analysis. This can be informed by observational studies. Relative treatment effects seen in randomised trials may then be applied to data on the baseline risk of health outcomes for the populations or subgroups of interest.” NICE manual 2022

“Specifically, the committee thought that using randomised data to estimate absolute event rates runs the risk of results that do not reflect NHS practice. It also thought that using observational data to estimate relative effects runs the risk of biased treatment effects because of unadjusted confounding variables. The committee noted that NICE’s technical support document 13 makes this distinction, advocating registry data to estimate absolute baseline event rates and randomised evidence to quantify relative differences. The committee concluded that it still preferred using the real-world evidence to estimate survival for people having cabazitaxel and the network meta- analysis to estimate the relative treatment effect of cabazitaxel compared with lutetium-177” ID3840 ACD2

4.3.5.1. Extrapolation of survival curves

Extrapolation of survival curves was conducted in accordance with NICE TSD 14 and NICE TSD 21. In order to determine if more flexible models were required log-cumulative hazard plots were examined to determine whether or not if they were not approximately straight lines. The company provided log cumulative hazard plots for OS and PFS in response to clarification question A1 for the ITT population and both risk subgroups. The survival analysis output from the R package for the UK RWE, CheckMate 9ER and CheckMate 025 is presented in Appendix K. There was no indication that more flexible models were required.

Standard parametric models were therefore fitted in line with TSD 14: exponential, Weibull, lognormal, log-logistic, Gompertz, gamma and generalised gamma using the flexsurvreg package in R.

The base case survival curve for each endpoint at each line and in each population was selected according to the following criteria which are listed in indicative priority order:

- Clinical validity – both in the biological plausibility of the trends in the hazard function considered via qualitative clinical input and in the absolute survival predicted versus quantitative clinical input from structured expert elicitation
- Consistency with longer term external data
- Consistency and validity across endpoints
  - Extrapolations where curves cross will be ruled out where possible
  - When using the PartSA approach the implications of selected OS and PFS curves on post progression survival and plausibility of this will be carefully considered
  - The overall modelled OS does not exceed the expected OS for the general population
- Statistical goodness of fit within trial (AIC and BIC) – curves with an AIC within 5 points of the best fitting curve are considered to have a similar goodness of fit
- Visual inspection
- Statistical validity versus the NMA type to be applied (the lognormal and loglogistic curves are not consistent with the application for a FP NMA) – this issue is acknowledged but was considered the lowest priority

This approach aligns with the guidance within TSD21: “careful thought should be given to the biological and clinical justification to any statistical approach selected; the approaches detailed herein should not be considered as an extended list of survival methods to “try out” on data. Instead, care should be taken to think through the underlying mechanisms likely to be dictating short and long-term hazard survival functions.”

Input from clinical experts was that the hazard function PFS would be expected to initially rise as those who are not sensitive to treatment progress early (first 1-2 years) followed by a slowing in the hazard function as those patients remaining are those who experienced initial disease control. In the longer term they would expected acquired resistance and general population mortality to take over with the potential for a late increase in hazards beyond the extent of current observed data. Given this curves which experienced continuing increase in hazards were ruled out as implausible.

Two datasets were identified which contained longer term data for sunitinib than CheckMate 9ER: CheckMate 214 and KeyNote 426. These datasets were used to assess consistency with longer term data.

Between one and three curves were selected for each endpoint to be tested in scenario analysis with the number selected based upon how similar the long-term projections were across curves.

In the maximum case a distribution with more pessimistic, more optimistic and similar (clone) projections was selected with attention paid to the same criteria as the base case in selection.

The next sections present the survival curve selections for each of the endpoints used within the state transition and PartSA scenarios for the reference curve for the 1st line all risk population in the model base case (sunitinib in the UK RWE). All other curve selections are presented in Appendix K.

Time to treatment discontinuation

Time on treatment was calculated in the base case using extrapolation of TTD curves where possible. A scenario analysis is included using PFS curves for all trials given the low level of reporting of TTD information across trials.

TTD information was only available for the UK RWE, CheckMate 9ER and CheckMate 214. Given this within the model base case we use TTD information from the reference curve for the UK RWE (base case) or CheckMate 9ER (scenario analysis) and apply the relative effects from the network meta-analysis of PFS. This is expected to provide a reasonable approximation for time on treatment given the close correlation between TTD and PFS observed in CheckMate 9ER and the UK RWE. The two exceptions to this were:

- Within CheckMate 214 the relationship between PFS and TTD differs for nivolumab + ipilimumab with PFS considerably longer than would be expected for the TTD observed – a simple scenario analysis has been carried out reducing TTD in line with the observed data using the estimated hazard ratio between PFS and TTD observed in the trial [redacted] acknowledging that proportional hazards may not hold this at least gives some indication of the expected scale of impact, the EAG does not have access to data on a per patient level which would allow more robust analyses to be carried out)
- Within CheckMate 9ER the observed TTD is slightly longer than the observed PFS for the cabozantinib + nivolumab arm. The impact of using data directly from CheckMate 9ER is tested in scenario analysis

Figure 43 shows that the data for TTD are mature within the UK RWE and that there is little difference in the curve fits. Table 63 shows the results of the curve fitting selection process. The log-logistic curve was selected within the model base case as this provided a good statistical and visual fit and had patients remaining on treatment after 6 years which is consistent with data from CheckMate 214 with the Weibull curve used in scenario analysis as a more pessimistic fit.

Figure 43: Curves fitted to TTD for suni in the UK RWE all risk population

Note: the number at risk is lower for TTD as a number of patients were excluded from the analysis due to invalid or unavailable treatment discontinuation times

Table 63: TTD curve selection for suni in the UK RWE all risk population

	Exp	Wei	Gomp	LogN	Logl	G	GG
Clinical validity hazards	✓	✓	✓	✓	✓	✓	✓
Consistency with CheckMate 214¥	✗	✗	✓	✓	✓	✗	✓
Statistical goodness of fit*	✗	✗	✗	✗	✓	✗	✗
Visual inspection	✗	✗	✓	✓	✓	✗	✓

Abbreviations: Exp, exponential; G, gamma, GG, generalised gamma; Gomp, Gompertz, int, intermediate; LogN, log normal; Logl, log logistic; PFS, progression free survival; RWE, real world evidence; suni, sunitinib; Wei, Weibull

Note: data were only collected for TTD at 1st line in the UK RWE dataset

*AIC within 5 of best fitting curve, underlined curve best statistical fit

¥ Some patients remained on treatment after 6 years

Stopping rules apply for a number of treatments for RCC. Where this is the case, data on the number of doses are used in preference to TTD data; where this has not been reported stopping rules will be applied after production of the expected TTD curve to calculate costs.

Table 64: Mean number of doses for treatments with a fixed duration

	Maximum duration	Mean number of administrations (SE)	Source
Nivo as part of cabo+nivo	2 years	[redacted]	Calculated from mean duration supplied by Ipsen of [redacted] months
Pem as part of pem+lenv	35 x 3 weekly cycles	12.3 (NR)	Calculated from mean duration of 17 months

Abbreviations: SE, standard error; NR, not reported

For combination therapies, in line with standard trial reporting, the TTD curve will only class patients as coming off treatment when both parts of the combination have been discontinued. We account for the reduction in drug cost with early discontinuation of one part of the combination using RDI data for each drug within the combination.

Treatment breaks are often used to allow toxicities to settle. NHSE restricts the length of treatment breaks before therapy is restarted, people who have longer breaks are not able to restart therapy via the normal funding route. Breaks of up to three months are allowed for nivolumab + ipilimumab and nivolumab monotherapy, 12 weeks for pembrolizumab + lenvatinib and avelumab + axitinib and 6 weeks for cabozantinib, tivozanib and lenvatinib + everolimus.[37] Similar restrictions are expected for other TKIs not included in the CDF drugs list. Treatment breaks will be considered within the model using RDI data to account for the impact on cost. The impact on effectiveness is assumed to already be included within the TTD data used to populate the model as people on a break will still be classed as remaining on treatment.

In practice, people are able to discontinue 1st line TKI monotherapy and switch to another TKI. This is only possible when they have had immediate prior treatment with a TKI which has had to be stopped solely as a consequence of dose-limiting toxicity and in the clear absence of disease progression.[37] This does not occur frequently (2.8% of patients switched TKI in the UK RWE) therefore these types of switches have been excluded from consideration within the decision model.

Progression free survival

Figure 44 shows that similar to TTD the Kaplan Meier curve for PFS is mature and there is little variation cross curve fits.

Figure 44: Curves fitted to PFS for suni in the UK RWE all risk population

Abbreviations: PFS, progression free survival; KM, Kaplan-Meier; RWE, real world evidence; UK, United Kingdom

The results from the expert elicitation exercise are presented in Appendix K. As noted in Section 4.2.5 for all of the reference curves considered the experts predictions at 3 years were above those in the observed data. The conditional survival probabilities between 3 and 5 years and between 5 and 10 years, were, however, consistent with a number of the potential models fitted to the observed data and these were used within the curve fitting process with a value within the 95% CI of estimates provided viewed as in-keeping with expert views.

Table 65 shows the results of the curve fitting selection process. The loglogistic curve was selected in the base case as this was consistent with available external data and the conditional survival probabilities from the expert elicitation in the individual risk groups. The Weibull curve was used in scenario analysis as a more pessimistic fit.

Table 65: PFS curve selection for suni in the UK RWE all risk population

	Exp	Wei	Gomp	LogN	Logl	G	GG
Clinical validity hazards	✓	✓	✓	✓	✓	✓	✓
Consistency with external data+	✗	✗	✓	✓	✓	✗	✓
Statistical goodness of fit*	✗	✗	✗	✗	✓	✗	✗
Visual inspection	✗	✗	✓	✓	✓	✗	✓

Abbreviations: Exp, exponential; G, gamma, GG, generalised gamma; Gomp, Gompertz, int, intermediate; LogN, log normal; Logl, log logistic; PFS, progression free survival; RWE, real world evidence; suni, sunitinib; Wei, Weibull

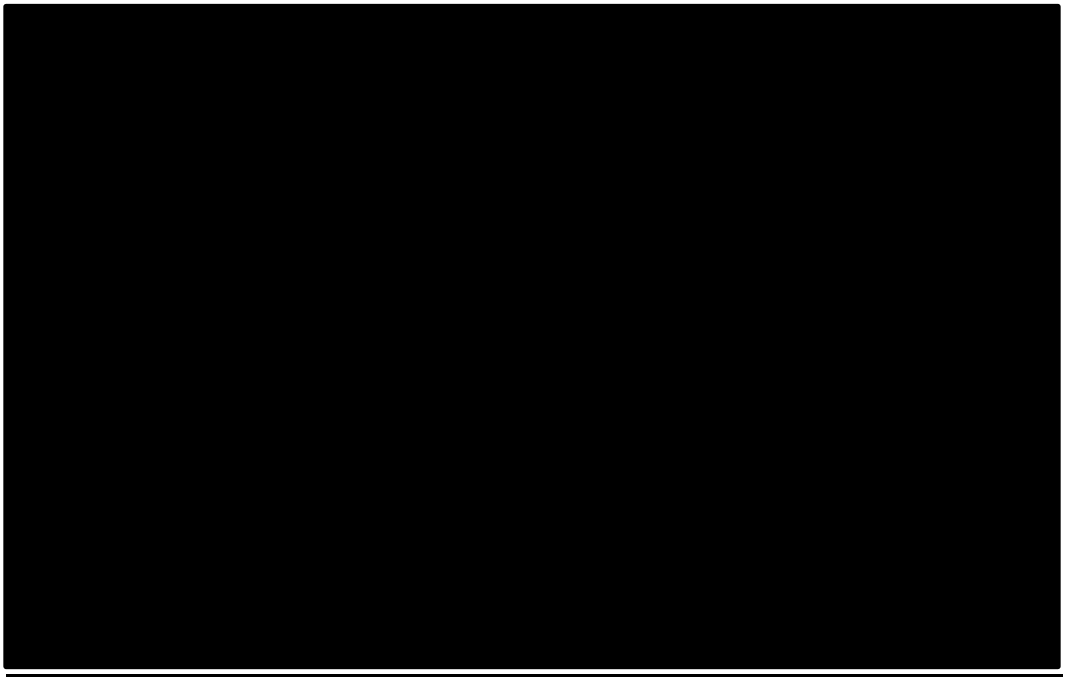
*AIC within 5 of best fitting curve, underlined curve best statistical fit

+ Given differences in populations included (RWE vs trials) curves were only ruled out if no patients remained in PFS at a timepoint clinical trial data (CheckMate 214 and KeyNote 426) indicated there should be patients remaining

Time to progression

Figure 45 shows that the TTP curve also has a high level of maturity. Table 66 shows the results of the curve fitting selection process. Consistent with TTD and PFS, the loglogistic curve was selected in the base case as this was consistent with available external data with the Weibull curve used in scenario analysis as a more pessimistic fit.

Figure 45: Curves fitted to TTP for suni in the UK RWE all risk population



Abbreviations: TTP, time to progression; KM, Kaplan-Meier; RWE, real world evidence; UK, United Kingdom

Table 66: TTP curve selection for suni in the UK RWE all risk population

	Exp	Wei	Gomp	LogN	Logl	G	GG
Clinical validity hazards	✓	✓	✓	✓	✓	✓	✓
Consistency with external data	NA						
Statistical goodness of fit*	✗	✗	✗	✗	✓	✗	✗
Visual inspection	✗	✗	✓	✓	✓	✗	✓

Abbreviations: Exp, exponential; G, gamma, GG, generalised gamma; Gomp, Gompertz, int, intermediate; LogN, log normal; Logl, log logistic; PFS, progression free survival; RWE, real world evidence; suni, sunitinib; Wei, Weibull

*AIC within 5 of best fitting curve, underlined curve best statistical fit

Overall survival (PartSA scenario analysis only)

There is more variation in the predictions using the extrapolated curves for OS than for the other endpoints as the data are less mature (Figure 46).

Figure 46: Curves fitted to OS for suni in the UK RWE all risk population

Abbreviations: OS, overall survival; KM, Kaplan-Meier; RWE, real world evidence; UK, United Kingdom

The loglogistic and lognormal curves both predict a much higher survival with a longer tail than the other curves in line with the nature of their underlying functions. These were not considered reasonable relative to the age of the patient population. All fitted curves except for the lognormal were considered to be of a similarly good statistical fit with all curves except the lognormal and

loglogistic curves also producing a good visual fit. The Gompertz was ruled out as the cumulative hazard function did not behave as expected. Given the similarity of the remaining curves the exponential was selected as the base case as the best statistical fit with the Weibull tested in scenario analysis as another plausible alternative.

Table 67: OS curve selection for suni in the UK RWE all risk population

	Exp	Wei	Gomp	LogN	Logl	G	GG
Clinical validity hazards	✓	✓	✗	✓	✓	✓	✓
Consistency with external data+	✓	✓	✓	✓	✓	✓	✓
Below general population?	✓	✓	✓	✗	✗	✓	✓
Statistical goodness of fit*	✓	✓	✓	✗	✓	✓	✓

Abbreviations: Exp, exponential; G, gamma, GG, generalised gamma; Gomp, Gompertz, int, intermediate; LogN, log normal; Logl, log logistic; PFS, progression free survival; RWE, real world evidence; suni, sunitinib; Wei, Weibull

*AIC within 5 of best fitting curve, underlined curve best statistical fit

+ Given differences in populations included (RWE vs trials) curves were only ruled out if no patients remained in OS at a timepoint clinical trial data indicated there should be patients remaining

Post progression survival

Within the state transition model up to three subsequent lines of treatment were allowed. The reference curve used for 2nd and 3rd line was cabozantinib. Results of curve fits to the endpoints of cabozantinib can be found in Appendix K. For 4th line the sample size was too small for a reference treatment to be selected within the dataset. For simplicity and given clinical expert advice that prognosis worsens as patients move down the lines, a Cox PH analysis was conducted using the UK RWE to determine the difference in outcomes between 3rd and 4th line which was then applied to all treatments equally to down-weight expected outcomes at 4th line relative to 3rd line (Table 68). This was done by 'stacking' 3rd line and 4th line survival times for patients and then estimating a hazard ratio with cluster-robust standard errors.

Table 68: Cox PH analysis comparing 3rd and 4th line outcomes in the UK RWE

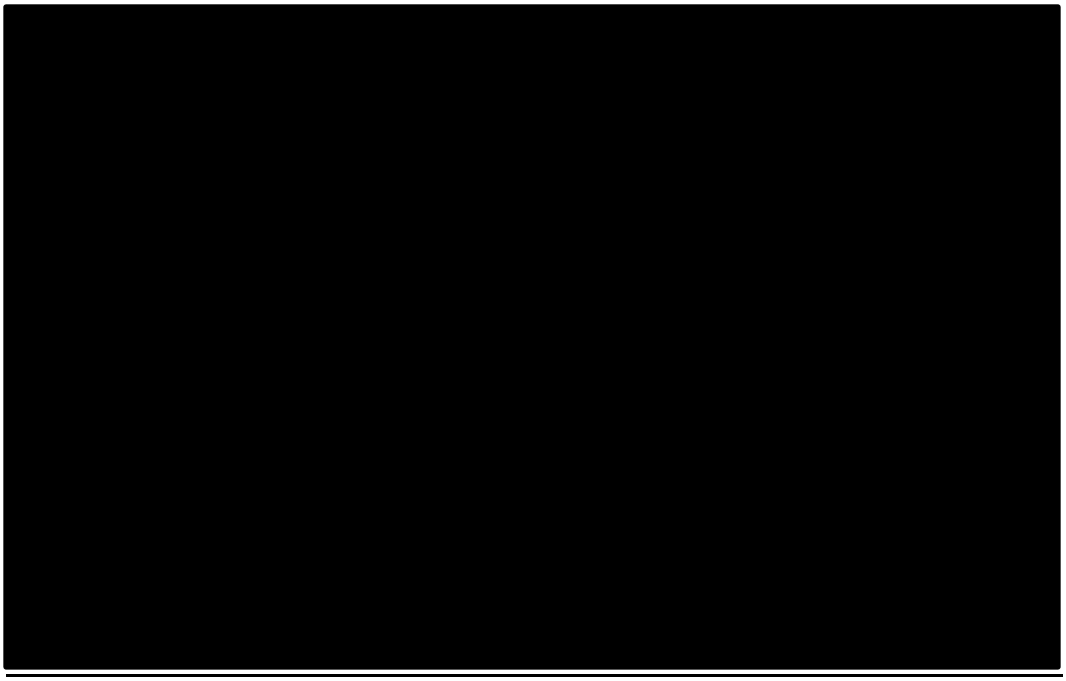
	Number of subjects / number of failures	Hazard ratio (95% CI)
OS	258 / 166	2.01 (1.45, 2.78)
PFS	237 / 176	1.74 (1.21, 2.51)

Abbreviations: OS, overall survival; PFS, progression free survival

For best supportive care pooled PPS outcomes for 4th line were taken for all patients (Figure 47). Outcomes are relatively uncertain as there were only 19 patients in the dataset, however,

the majority of patients experienced outcomes early within the dataset in line with clinical expert advice.

Figure 47: Curves fitted to PPS for 4L patients in the UK RWE all risk population



Abbreviations: RWE, real world evidence; UK, United Kingdom

The lognormal curve was selected as the most appropriate for BSC based on consistency the conditional survival probabilities from the expert elicitation exercise, it should be noted, however, that there is little difference between the fitted curves due to the maturity of the data (Table 69). The exponential curve was tested in scenario analysis as a more pessimistic option and the best statistical fit.

Table 69: BSC curve selection in the UK RWE

	Exp	Wei	Gomp	LogN	Logl	G	GG
Clinical validity hazards	✓	✗	✓	✓	✓	✗	✓
Consistency expert elicitation	✓	✓	~	~	✓	✓	✓
Consistency with external data+	NA						
Statistical goodness of fit*	✓	✓	✓	✓	✓	✓	✓
Visual inspection	✓	✓	✓	✓	✓	✓	✓

Abbreviations: BSC, best supportive care; Exp, exponential; G, gamma; GG, generalised gamma; Gomp, Gompertz; LogN, log normal; Logl, loglogistic; RWE, real world evidence; UK, United Kingdom; Wei, Weibull

*AIC within 5 of best fitting curve, underlined curve best statistical fit

+ Expert elicitation values for evero at 4th line: mean 25.1% at 3 years (17%, 34%) conditional survival between 3 and 5 years 36.3% (19%, 52%) conditional survival between 5 and 10 years 37.2% (13%, 56%)

Final selected curves

Table 70: Final selected curves for suni using the UK RWE

	All risk population		Int/poor risk population		Favourable risk population	
	Curve selection	**Rationale**	**Curve selection**	**Rationale**	**Curve selection**	**Rationale**
TTD	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. Consistent with CheckMate 214 data Consistent with PFS	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Consistent with PFS Consistent with all risk population	Base case: loglogistic Scenarios: generalised gamma	Good statistical and visual fit. All curves provide similar AUC Consistent with PFS Consistent with all risk population
PFS	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. Broadly consistent with external data	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Broadly consistent with external data and expert elicitation Consistent with all risk population	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Broadly consistent with external data and expert elicitation Consistent with all risk population
TTP	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. Consistency with PFS selection	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Consistency with PFS selection Consistent with all risk population	Base case: loglogistic Scenarios: Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Consistency with PFS selection Consistent with all risk population
OS	Base case: exponential Scenarios: Weibull	Good statistical and visual fit Midrange estimate within plausible curves	Base case: exponential Scenarios: Weibull	Good statistical and visual fit Consistent with all risk population	Base case: Exponential Scenarios: Weibull	Good statistical and visual fit Midrange estimate Consistent with all risk population

	All risk population		Int/poor risk population		Favourable risk population	
	Curve selection	**Rationale**	**Curve selection**	**Rationale**	**Curve selection**	**Rationale**
PPS	Base case: lognormal Scenarios: exponential	All curves similar AUC due to completeness of KM Most consistent with expert elicitation Note Kaplan Meier based on 19 patients	NA	NA	NA	NA

Abbreviations: AUC, area under the curve; KM, Kaplan, Meier; NA, not applicable; OS, overall survival; PFS, progression free survival; TA, technology appraisal

4.3.5.2. Calculation of relative treatment effectiveness

Treatment effectiveness for all other therapies has been calculated by applying the results of the NMAs conducted by the EAG in the base case. In scenario analysis we explore the impact of using individually fitted curves to the cabozantinib + nivolumab trial data when using the trial only scenario analysis.

Table 71 provides a summary of where relative effectiveness has been taken from for each of treatments for each endpoint. For first line treatments in the model base case the FP NMA is used where this is available except in the case of pem+lenv where the FP NMA produced implausible results; moreover, PFS curves in intermediate/poor risk are not available for this treatment.. It is acknowledged that use of the PH NMA will bias towards pem+lenv as the CLEAR trial demonstrated non-proportional hazards (curves coming together), the extent of bias is, however, expected to be mitigated by the application of treatment-effectiveness waning in the model base case. For 2nd line and 3rd line treatments we use the PH NMA in preference to the FP NMA due to the sparsity of the available network and extreme results within the fitted models, and our view that the PH NMA likely reflects a more reliable estimate of relative effectiveness. We assume equivalence of sunitinib, pazopanib and tivozanib in the model base case as none of these treatments were available in the FP NMA and tivozanib was not available for OS in the PH NMA. This is in line with prior appraisals which concluded that:

- Pazopanib and sunitinib have similar effectiveness (TA858, TA645)
- Tivozanib is at best similar to pazopanib and sunitinib (TA858, TA645)

In the base case we use the NMA results for everolimus and axitinib, we tested in scenario the assumption that everolimus and axitinib have similar effectiveness (TA432, TA417).

Table 71: Base case application of relative effectiveness in the economic model

	TTD	**PFS**	**TTP**	**OS**
1L				
Cabo+nivo	Rel. effect = PFS	FP NMA	Rel. effect = PFS	FP NMA
Nivo+ipi	Rel. effect = PFS	FP NMA	Rel. effect = PFS	FP NMA
Pem+lenv	Rel. effect = PFS	PH NMA¥	Rel. effect = PFS	PH NMA¥
Ave+axi	Rel. effect = PFS	FP NMA	Rel. effect = PFS	PH NMA
Suni	Reference	Reference	Reference	Reference
Pazo	Equal to suni*	Equal to suni+	Equal to suni*	Equal to suni+

	TTD	**PFS**	**TTP**	**OS**
Tivo	Equal to suni*	Equal to suni+	Equal to suni*	Equal to suni*
Cabo	Rel. effect = PFS	FP NMA	Rel. effect = PFS	FP NMA
2L& 3L				
Nivo	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA
Pazo	HR to PFS	Equal to tivo*	Rel. effect = PFS	Equal to tivo*
Tivo	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA
Suni	HR to PFS	Equal to tivo*	Rel. effect = PFS	Equal to tivo*
Cabo	HR to PFS	Reference	Reference	Reference
Lenv+evero	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA
Evero	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA
Axi	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA

Abbreviations: HR, hazard ratio; NMA, network meta-analysis; OS, overall survival; PFS, progression free survival; PH, proportional hazards; TTD, time to discontinuation; TTP, time to progression; rel. effect; relative effectiveness

*Data not available in either NMA

+ PH NMA available but not used in base case

¥ FP NMA only available for the all risk population for PFS, PH NMA used due to the FP NMA producing implausible results, this is likely to bias towards pem+lenv

For TTD and TTP where we do not have NMAs conducted due to the sparsity of data in the base case we assume that the PFS hazard ratio for 1st line applies to TTD and TTP as discussed previously. We use the same method for TTP at 2nd and 3rd line. For later lines for TTD as data were not available in the UK RWE we use the hazard ratio between TTD and PFS calculated at 1st line for all treatments:

- TTD HR to PFS: 1.19 (1.15, 1.24)

For 4th line outcomes we apply the hazard ratio between pooled 3rd and 4th line outcomes calculated from the UK RWE to all treatments and then calculate TTP based upon its relationship to PFS at earlier lines.

- 4th line OS HR 2.01 (1.45, 2.78)
- 4th line PFS HR 1.74 (1.21, 2.51)
- TTP HR to PFS: 0.82 (0.80, 0.84)

4.3.5.3. Treatment effectiveness waning

Following application of NMA results we considered the plausibility of the long-term treatment effect predicted for each of the treatments relative to the reference treatment. The application of treatment effect waning assumptions for IO/TKI and IO/IO combinations was considered for each treatment based upon:

- How long the treatment is given for
- The mechanism of action of the treatment and biological plausibility informed by clinical expert advice
- The trends seen within the trials (Figures 30 and 31) and the fitted FP NMA models (see Section 3.7.3)
- Consistency between treatments with similar mechanisms of action
- Precedent in prior appraisals

Precedent was used to guide considerations. Table 72 demonstrates that within RCC, as in many other oncology indications, Committee concerns regarding uncertainty in long-term treatment effects in earlier submissions led to modelling of scenarios around TE waning in later submissions and assumptions becoming part of the base case where stopping rules for treatments were in place, follow-up was particularly short or OS curves crossed. We would note, however, that even in TA858 where follow-up was longer and stopping rules did not apply the Committee considered exclusion of TE waning from the EAG base case to be uncertain.

Looking firstly at cabozantinib + nivolumab the hazard plots supplied by Ipsen in response to clarification questions A21 (44-month datacut) indicate that [redacted]. A similar trend is not seen for PFS.

When looking at the information available across IO / TKI combinations (Figures 30 and 31) the longest-term data available is for pembrolizumab + axitinib (median 67.2 months) which is not recommended in England. Here a clear trend can be seen for OS of increasing hazard ratios (hazard ratios getting closer to 1) with later datacuts and the OS Kaplan Meier appears to be starting to converge with the sunitinib arm at the latest times (acknowledging relatively low numbers at risk). A similar pattern of increasing OS hazard ratios and convergence of Kaplan Meier's can be seen over time for pembrolizumab + lenvatinib for which the latest datacut has a median follow-up of 49.8 months. For PFS the same convergence cannot be seen in the pembrolizumab + axitinib data. In the pembrolizumab + lenvatinib data there is some indicates

of the curves starting to converge and the HR per datacut has seen a small increase over time for cabozantinib + nivolumab (0.51 to 0.59 from first to latest datacut) and pembrolizumab + lenvatinib (0.41 to 0.47).

For nivolumab + ipilimumab there is no clear trend in the HRs by datacut for either OS or PFS and there is no evidence of Kaplan Meier curves coming together for either OS or PFS in the latest datacut (67.7 months).

Input from clinical experts was that IO / TKI combinations would be expected to act similarly in terms of the durability of long-term relative effectiveness compared to TKI monotherapy.

A recent podcast[229] following considerable discussion regarding the latest results released at ASCO summarises well the lack of agreement within the clinical community on the long-term effectiveness of IO/TKI combinations. There are essentially two schools of thought:

- The OS curves coming together is expected and similar to what was observed for IO/BRAF combinations in melanoma. This could be due to initial responses being TKI driven, benefit being lost when TKIs are stopped and/or combining IOs and TKIs being unhelpful in terms of getting the best immune response due to the toxicity of the TKI component precenting the best results being achieved by the IO component
- The OS curves coming together is an artefact of low numbers at risk.

One thing is clear, the most recent datacuts have added to, rather than reduced, uncertainty regarding the long-term effectiveness of IO / TKI combinations.

Our FP NMA shows that with the models selected for the base case there is an upward trend in the hazard ratios for the IO / TKI combinations for OS. This is not the case for PFS with the exception of pembrolizumab + lenvatinib.

All of the IO / TKI combinations in the decision problem for cabozantinib + nivolumab have a stopping rule in place for the IO component, whereas there is no stopping rule in place for nivolumab maintenance within the nivolumab + ipilimumab component.

Given that stopping rules are in place and more mature datacuts have added uncertainty to the durability of the long -term effect for IO / TKIs the EAG base case applies treatment effect waning at 5 years to all IO / TKI combinations based on hazards, all endpoints. Five years was selected as the longest timepoint at which data is available for 1st line combinations with a

reasonable number at risk remaining. IO / TKI combinations are assumed to wane towards the reference curve (sunitinib).

The following scenarios are tested within the EAG analysis:

- Waning applied at 10 years to all IO / TKI combinations based on hazards, all endpoints
- Waning applied at 10 years to all IO combinations based on hazards, all endpoints
- Waning applied between five and 20 years to all IO / TKI combinations based on hazards, all endpoints
- Waning applied between five and 20 years to all IO combinations based on hazards, all endpoints
- No treatment effect waning

These scenarios are all more optimistic than the base case due to the maturity of the available data and difficulties modelling a direct impact on OS in a state transition framework where OS is driven instead by the mix of subsequent therapies.

The following additional scenarios are applied when presenting the PartSA:

- Waning applied to OS only at five years to all IO / TKI combinations based on hazards
- Pessimistic scenario: waning applied between four and six years to all IO/TKI combinations based on absolute survival for OS only, this is based on the timing of convergence of the OS curves for pembrolizumab + lenvatinib and pembrolizumab + axitinib.

The latter scenario represents the worst-case scenario if the fears around IO/TKI lack of long-term durability of effect discussed at ASCO 2023 play out.

Treatment effect waning has not been applied for 2nd line and later treatments as mature data exists for CheckMate 025 (median 87.7 months) where there is no indication of convergence of the Kaplan Meier curves and the majority of other treatments included in the network have the same mechanism of action as the reference treatment.

In order to avoid implausible results in cases where the hazards were higher with the intervention prior to the application of treatment effect waning we retain the original hazards rather than lowering them to match the reference curve.

Table 72: Precedent from prior appraisals on treatment effect waning

TA	Treatment type	Stopping rule prior to progression?	OS follow-up	Committee considerations on TE waning
TA858	IO+TKI	No	Median 33 months	Excluded from EAG base case, Committee considered uncertain
TA780	IO+IO	Ipi only given during first 4 cycles	Min 60 months	Death hazards between arms would be likely to equalise, probably between 4.5 and 21 years
TA650	IO+TKI	Yes	Median 13 months	5 year TE waning (also looked at 3 and 10 years) regardless of response
TA645	IO+TKI	No	Min 13 months	Excluded after removal of stopping rule, Committee request presented TE over time
TA542	TKI	No	Median 29 months OS curves crossed	Modelling should assume that there is no treatment effect beyond the observed survival data, which covered a duration of less than 4 years. EAG base case 5 year TE waning accepted
TA498	TKI+mTOR	No	> Median 25 months*	Lifetime treatment effect in EAG base case. Committee would have liked to have seen more conservative assumptions explored
TA463	TKI	No	Median 21 months	Assuming the effect of cabo continues for up to 30 years, based on a trial with a median follow-up of under 2 years for overall survival, was highly uncertain
TA417	IO	No	Median 17 – 18 months	Committee remained concerned that the company assumed a continual post-treatment benefit of nivo and had not presented to the Committee analyses that excluded this benefit

Abbreviations: EAG, external assessment group; IO, immunotherapy; OS, overall survival; mTOR, mammalian target of rapamycin inhibitor; TE, treatment effect; TKI, tyrosine kinase inhibitor

Notes:
*Follow-up only reported for Dec 2014 data-cut, July 2015 data-cut used in model

4.3.5.4. Accounting for general population mortality

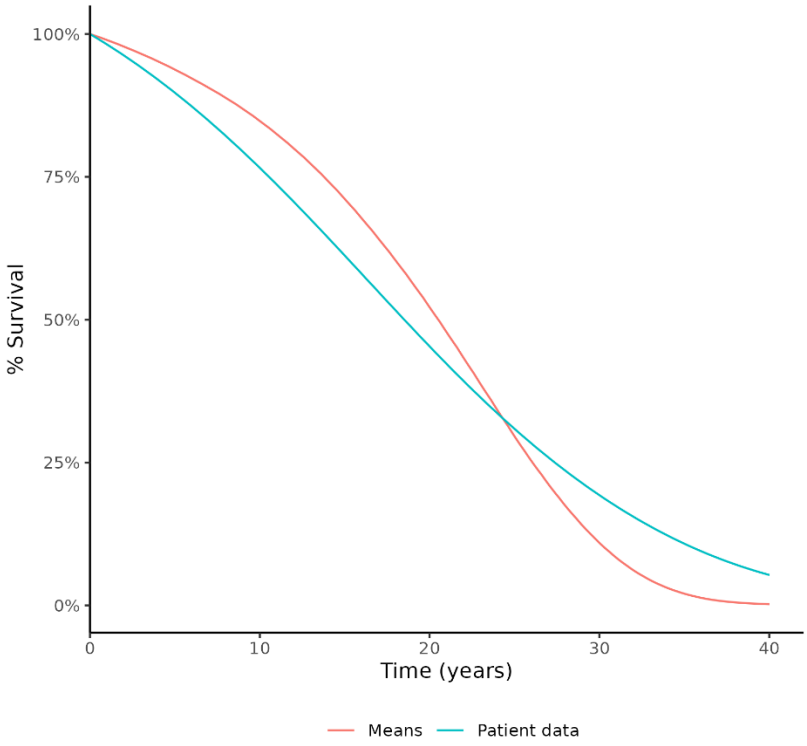
In addition to the base check that the predicted survivor function for OS does not exceed that of the general population we ensure that the hazard function for OS does not fall below that of the general population for any of the modelled cycles.

As the EAG does not have access to cause-specific death data survival curves we have used a simple method (selection of the maximum hazard function for any time period) to account for any issue of patients with RCC being projected to live longer than those in the general population with the same age and sex mix at baseline. Other alternatives such as the relative survival models described in TSD21 require cause specific mortality data.

ONS life tables[230] were used to calculate mortality for the general population with age and sex data for patients at the start of treatment taken from UK RWE if possible. Data were used from 2017-2019 as 2018-2020 values were affected by COVID. We model mortality separately by sex accounting for the differences in life expectancy by gender.

Figure 48 shows the expected general population mortality for people with an age and sex profile matching the 1st line all risk population in the UK RWE. This demonstrates that a maximum time horizon of 40 years is appropriate and the difference that the method for calculation of general population mortality makes. Using the full age and sex demographics produces a steeper drop at the beginning of the curve and a longer tail than assuming all patients have the same mean age.

Figure 48: Expected general population survival: age and sex matched to the UK RWE



4.3.5.5. Adjustment for curves crossing

Whilst every effort has been made to ensure that curves do not cross during survival curve selection this may be unavoidable for outcomes where curves are close together (e.g. TTP and PFS). In these cases, we adjust curves such that PFS <= TTP and PFS <= OS to remove any logical inconsistency. We had initially considered applying a restriction that TTD <= PFS, however, as some patients in the dataset continued to receive treatment beyond progression this was not considered appropriate.

4.3.5.6. Calculation of final outcomes by first line treatment

Within the state transition analysis first the survival curves are calculated for each treatment available in practice at each line included within the model. Health state occupancy is then calculated for each possible treatment sequence. Possible treatment sequences were defined by the following rules which were tested with clinical experts (see Appendix M):

- Ave+axi1L in any risk
- Cabo+nivo 1L in any risk

- Suni 1L in any risk
- Pazo 1L in any risk
- Tivo 1L in any risk
- Nivo+ipi 1L in intermediate/poor risk only
- Pem+lenv1L in intermediate/poor risk only
- Cabo 1L in intermediate/poor risk only
- Nivo+ipi, pem+lenv, ave+axi, cabo+nivo and nivo cannot be used if an IO was used in the last 12 months in the adjuvant setting
- Only one of nivo+ipi, pem+lenv, ave+axi, cabo+nivo and nivo within the treatment pathway
- Axi, cabo, lenv+evero, suni, tivo, evero, pazo, nivo can all be used 2nd and 3rd line
- Axi and evero can be used 4th line
- Lenv+evero can only be used after one prior anti-VEGF (ave+axi, axi, cabo cabo+nivo, pazo, pem+lenv, suni, tivo)
- Suni, tivo and pazo when 2L+ can only be used after nivo+ipi, pem+lenv, ave+axi and cabo+nivo
- The same treatment cannot be used twice (either as monotherapy or as part of a combination)

Once health state occupancy was calculated for each treatment sequence the expected outcomes given the first-line treatment were calculated by weighting each possible sequence by the percentage of patients expected to receive that sequence (see Section 4.3.8.6). In the base case this was informed by the UK RWE, in scenario analysis use of trial data is tested.

4.3.5.7. Validation

Within the model results and validation addendum which will follow this report we will present the final modelled curves vs Kaplan Meier data and compare outcomes for the restricted mean survival time, including for OS, based upon the aggregation of outcomes for each line of treatment to determine whether the model fit is appropriate. The model curve will then be compared to the projections from other models previously used for NICE STAs in the same decision point.

4.3.5.8. Exploratory analysis looking at the impact of prior adjuvant therapy

Based upon the information provided during expert elicitation the impact of prior adjuvant therapy is expected to be different according to the type of treatment with prior adjuvant therapy expected to negatively impact on outcomes for cabozantinib + nivolumab even after a wait of at

least a year in line with NHS criteria and expected to positively impact on outcomes with sunitinib (as patients who receive adjuvant therapy are scanned more frequently and therefore disease progression is expected to be picked up at an earlier stage). The EAG conducted an exploratory analysis looking at the impact of prior adjuvant treatment based upon the outcomes of the expert elicitation exercise, acknowledging that the number of experts who answered these questions was low (n=2 or 3). This analysis compared the expected survival at the 3-, 5- and 10-year timepoints for each treatment using information from the experts who answered the questions related to adjuvant treatment only. The average hazard ratio across the 3 timepoints available for sunitinib was 0.51 and for cabozantinib plus nivolumab was 1.36 accounting for the conditional survival format of the 5- and 10-year timepoints.

4.3.6. Adverse events

The impact of toxicity on both costs and quality of life has been included within the economic analysis. The impact of toxicity on discontinuation has been addressed through the TTD endpoint and not separately of other types of discontinuation given the data available.

Adverse events rates were taken from data supplied by Ipsen for CheckMate 9ER. The initial data request asked for these to account for cases where there are multiple events rather than just being the number of people experiences a specific type of adverse event. This was not supplied and adverse events were instead presented as is commonly the base according to the number of patients experiencing each type of event. This is not considered to be a major limitation.

The model included G3+ AEs which occur in more than 5% of patients in any trial arm in the model. This aligns with TA858.[38] In addition the following three adverse events were included at any grade on the advice of clinical experts that these were the AEs with most impact on patient quality of life and NHS resources at lower grades:

- Hand foot syndrome
- Diarrhoea
- Fatigue

All three of these were noted as common chronic VEGF toxicities with a large impact on patients.

Reporting of specific adverse events was inconsistent across the literature and producing NMAs per specific AE, given the number of interest, was not considered feasible therefore the following options are presented to capture the impact of toxicity within the model:

5. Base case: NMA relative effects applied to reference treatment (sunitinib (1st line) and everolimus (2nd line-plus)) and trial (CheckMate 9ER[59] and CheckMate025[89]) using EAG NMA for grade 3+ AEs and all grade NMA from the cochrane review[150] for the 3 specified Grade 1-2 AEs namely diarrhoea, fatigue and palmar-plantar erythrodysaesthesia syndrome
6. Scenario analysis: treatment related naïve AE rates for Grade 3+ (in ≥5% of patients) AEs (absolute estimates) from CheckMate 9ER or comparator pivotal trials – this is standard practice in the majority of oncology TAs

No data was available for adverse events from UK RWE for RCC specifically. One publication was identified focussing on safety outcomes for IOs which showed that from 2,125 patient records one third of patients experienced a clinically significant (Grade 3+) immune-related AE.[131] Real-world data from Germany indicated that 32/67 (48%) of patients receiving nivolumab + cabozantinib experienced Grade 3+ AEs.

AE rates per patient per cycle was calculated as: number of patients experiencing any grade or grade 3+ AEs/patient weeks observed (number of patients in the trial multiplied by the treatment duration in the trial). This is likely to underestimate the impact, however, data on the number of events experienced was not available.

AEs may either be applied as a per cycle event rate or as a one-off cost and utility impact at the start of each treatment. Given clinical advice that the majority of AEs occur within the first 6 months the model base case applies impact as a one-off. This is consistent with TA858.

In scenario analysis events were applied per cycle which assumes they are equally likely to occur for the entire duration of treatment as data was not available for the majority of treatments on when AEs occurred. Clinical expert advice was that IO-related toxicities are usually experienced within the first 6 months although late events can occur (but are rarely of major impact) and that TKI-related toxicities are also usually first experienced within the first six months but that cumulative fatigue is a major issue which continues into the longer-term.

These approaches are considered to give a reasonable approximation given that adverse events were not found to be a key model driver in any of the published literature.

The final costs and quality of life impacts for each treatment will be checked with clinical experts to ensure they hold face validity, if the experts indicate issues then scenarios provided by the experts will be considered.

Table 74 presents the rate per patient per week for the reference treatment (sunitinib) and Table 75 presents the relative risk estimates for comparators from the EAG NMA and Cochrane review.

Based on clinical expert advice that the impacts of diarrhoea are different dependent on whether it is IO or TKI induced the rates were split up for this specific adverse event. The rates were split up into IO or TKI induced based on the CheckMate 9ER data (Table 11 of the company evidence submission v2.0 dated 13042023[110]) which indicated 8 G3+ diarrhoea events were considered to be immune-mediated out of 28 events in total and 10 G1/2 diarrhoea events were considered to be immune-mediated related out of 182 events in total. It was assumed that same proportions apply to all IO/TKI combinations, for nivo+ipi and nivo monotherapy all diarrhoea events were 100% IO related and for all other treatments 100% TKI related, as mentioned in the Table 73 below.

Table 73. Diarrhoea events that are IO or TKI related for all treatments

Treatments	Diarrhoea (G3+)		Diarrhoea (G1/2)		Source/Assumption
	IO related (%)	TKI related (%)	IO related (%)	TKI related (%)	
Nivo	100%	0%	100%	0%	Assumed IO related
Cabo+nivo	[redacted]	[redacted]	[redacted]	[redacted]	CheckMate 9ER (company submitted data[110])
Nivo+ipi	100%	0%	100%	0%	Assumed IO related
Lenv+pem	29%	71%	5%	95%	Assumed same as cabo+nivo
Ave+axi	29%	71%	5%	95%	
Pazo	0%	100%	0%	100%	Assumed TKI related
Tivo	0%	100%	0%	100%	Assumed TKI related
Suni	0%	100%	0%	100%	Assumed TKI related
Cabo	0%	100%	0%	100%	Assumed TKI related
Lenv+evero	0%	100%	0%	100%	Assumed TKI related
Evero	0%	100%	0%	100%	Assumed TKI related
Axi	0%	100%	0%	100%	Assumed TKI related

Table 74: Adverse event rates per patient per week (reference treatment)

Adverse events	Suni 1L reference treatment	Evero 2L reference treatment
Grade 3+		
ALT increased	0.0000	0.0000
Anaemia	0.0000	0.0049
Decreased appetite	0.0000	0.0000
Diarrhoea	0.0023	0.0008
Fatigue	0.0009	0.0017
HFS or palmar-plantar syndrome	0.0021	0.0000
Hypertension	0.0031	0.0000
Hypertriglyceridemia	0.0000	0.0031
Hyponatraemia	0.0010	0.0000
Hypophosphatemia	0.0010	0.0000
Increase in lipase	0.0000	0.0000
Increased AST	0.0000	0.0000
Leukopenia	0.0000	0.0000
Lymphopenia	0.0000	0.0000
Nausea	0.0000	0.0000
Neutropenia	0.0000	0.0000
Platelets count decreased	0.0000	0.0000
Proteinuria	0.0000	0.0000
Stomatitis	0.0000	0.0000
Vomiting	0.0000	0.0000
Weight loss	0.0000	0.0000
Specified grade 1/2		
Diarrhoea	0.0107	0.0022
Fatigue	0.0083	0.0345
HFS or palmar-plantar syndrome	0.0087	0.0000

Abbreviations: HFS, Hand-foot syndrome

Table 75. Relative risk estimates (for AEs) from NMA

Treatments	Grade 3+	Source	Specified grade 1/2			
			Diarrhoea	Fatigue	HFS	Source
Sora	0.944	EAG NMA (1L)	1.95	0.62	4.80	Cochrane review (for nivo+ipi within trial relative risk from CheckMate 214 has been used as it is not available in the Cochrane review)
Cabo+nivo	1.238		1.57	0.73	1.00	
Nivo+ipi	0.808		0.52	0.86	0.03	
Lenv+pem	1.316		1.82	0.97	1.04	
Ave+axi	1.082		2.44	0.95	1.33	
Pazo	1.034		1.14	0.63	0.48	
Tivo	0.77		0.60	1.36	0.66	
Cabo	1.134 (1L)		0.92	0.38	1.85	
	1.367 (2L+)	EAG NMA (2L+)				
Lenv+evero	1.601		2.18	1.72	0.74	
Evero	1 (2L+)		0.18	1.79	0.10	
Axi	2.303		3.76	3.76	2.27	
Nivo	0.582		1	0.5	0	

Abbreviations: HFS, Hand-foot syndrome; NMA, network meta-analysis

Note: the Cochrane review assumes that the impact of cabozantinib on AEs is the same across lines of treatment

4.3.7. Utility values

4.3.7.1. Utility values from CheckMate 9ER

HRQoL data were collected in the CheckMate 9ER study using patient-reported outcome (PRO) instruments, including the EQ-5D-3L, EQ-VAS and the National Comprehensive Cancer Network (NCCN) Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI-19). The company provided the EAG with updated CheckMate 9ER HRQoL data on the 9th of May 2023. Based on this analysis, HRQoL data were available up to week 223, reflecting a longer timeframe than that reported by Cella et al.[122] (2022; median follow-up 23.5 months), which reported change in patient HRQoL from baseline to week 115. The number of patients included in the cabozantinib + nivolumab arm was reported to be n=320 and the number of patients in the sunitinib arm was n=319. EQ-5D data were not published for the most recent datacut at the time of writing.

For patients in the cabozantinib + nivolumab arm EQ-5D-3L data were collected on Day 1 of Week 1 of each 2-week study cycle and at the first two safety follow up visits (approximately 30 days and 100 days after the last nivolumab dose). For sunitinib patients EQ-5D-3L data were collected on Day 1 of Week 1 of each 6-week study cycle and at the first two safety follow up visits (approximately 30 days and 100 days after the last sunitinib dose). The EAG note that the estimation of utility values based on two data points, after stopping treatment with nivolumab (in the cabozantinib + nivolumab arm) and sunitinib introduces uncertainty into the analysis. This uncertainty is further compounded in the cabozantinib + nivolumab arm due to the 24-month stopping rule in place for nivolumab.

Overall, the EQ-5D-3L completion rate within the trial was considered reasonably high (88%). At baseline, 94% and 97% of patients in the cabozantinib + nivolumab arm and the sunitinib arm had completed the EQ-5D-3L respectively. Completion rates across treatment arms (and according to progression status) varied over time. The EAG noted that in the cabozantinib + nivolumab arm there was a marked increase in missing/not completed EQ-5D-3L data from week 179 to week 221, particularly for progressed disease patients. Further information regarding number of patients completing the EQ-5D-3L by health state can be found in Appendix G.

In their analysis of HRQoL data, the company used a mixed model repeated measures (MMRM) approach which included fixed-effect variables i.e. baseline EQ-5D-3L, week number of the visit

and adverse events. Random effects variables included, week number of the visit, adverse events, progression status, and prognostic status. The company's mixed model equation is outlined in Table 76 (for Visit i under patient j).

The company justified the use of a MMRM approach as the same patient needed to complete the questionnaire multiple times throughout the study period and a MMRM accounted for the hierarchical nesting of the data, which allowed for consideration of evolving intra-individual values, longitudinally, thus leading to more robust utility estimates. Whilst the EAG considered the use of a MMRM model to be reasonable, there was some uncertainty surrounding the company's approach to imputing missing values. During clarification the company was asked to comment on why the imputation was used and the exact methods applied. Based on their clarification response, imputation was conducted as some patients did not complete EQ-5D-3L questionnaires at follow up visits, which could introduce statistical bias, exaggerated type 1 error or reduced power. A single mean imputation was not undertaken as this would ignore the nature of hierarchically organised data. Furthermore, the company provided utility values based on a model without imputed estimates. The EAG noted that the utility values estimated without imputed estimates broadly aligned with the utilities based on modelled imputed estimates. The EAG considered the company's approach to be reasonable and noted that the use of imputed estimates did not appear to bias the analysis.

The estimates from the final model predicting EQ-5D-3L change from baseline are outlined in Table 76. The EAG noted several concerns surrounding the company's MMRM approach which include the following.

- Validity of the stepwise backward elimination method for model selection is unclear. Based on the EQ-5D-3L data provided to the EAG on the 9th of May, the company generated 12 models used to predict change in EQ-5D-3L from baseline, each with different fixed and random effects parameters. Based on the company's response to EAG clarification questions, the final model (used to estimate health state utilities) was selected based on the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The initial model contained a relatively large number of covariates including age, sex, race, first measurement of EQ-5D utility value, treatment group, adverse events, weeks of visit, progression status and prognostic score group. As part of the stepwise backward elimination approach, covariates were removed one by one. Once a covariate was removed, the model was compared to the previous best fitting model. If the model had a lower AIC/BIC than the previous best fitting model, the poorer fitting model was eliminated. Based on this method, the final model selected by the company did not include age, sex, race or treatment as covariates. Based on 'Model 6' provided by the company, age and treatment did not appear to be key determinants in the variability of EQ-5D-3L, suggesting that their exclusion may be reasonable.

- The EAG noted that cross/external validation of the stepwise backward elimination method was not discussed by the company. Other potential limitations including the sensitivity to the order in which the variables were removed were not highlighted. Overall, the EAG considered the methodological rigour of the approach as a means of model selection is associated with uncertainty and the level of uncertainty was not adequately categorised by the company. Furthermore, based on AIC/BIC statistics presented, several models could be considered broadly similar i.e. with less than five deviations in AIC/BIC between them. The company's decision to therefore select the model with the lowest AIC/BIC, whilst rational, is associated with uncertainty as other models could be considered reasonable. Ultimately, for each model generated, the company did not present health state utilities (based on progression status). Therefore, it was not possible to comment on the comparative validity of each model with respect to their generated health state utility values.
- The company provided detail on the MMRM approach used to estimate change in EQ-5D-3L from baseline and also provided summary statistic tables outlining utility by progression status and prognostic status, however the interim step detailing the calculations used to estimate the precise mean utilities was not provided. The company's clarification response to the EAG provided a description of the approach undertaken, however the granular calculations for utility estimation were not provided. This remains an area of uncertainty. The EAGs interpretation of the response provided is that the company use only the week numbers observed within the trial within the prediction of utilities. This is likely to overestimate the utility associated with the entire modelled horizon.

Table 76: Mixed model equation used by the company

Full mixed model equation	Yij = 0.008971 + 0.000003703*Week number of the visit ij+0.01065* AE ij+0.007209* Progression status ij+ 0.002978 * Prognostic status ij+ 0.3884+ (-0.49670) * First measurement of EQ5D-3L index value ij+ (-0.03339) * AE ij + (-0.00021) * Week number of the visit ij + 0.01276
Random effects	0.008971 + 0.000003703*Week number of the visit ij+0.01065* AE ij+0.007209* Progression status ij+ 0.002978 * Prognostic status ij+ 0.3884+
Fixed effects	(-0.49670) * First measurement of EQ5D-3L index value ij+ (-0.03339) * AE ij + (-0.00021) * Week number of the visit ij
Level-1 error variance	+ 0.01276

Utility values estimated by the company from CheckMate 9ER using the MMRM approach are outlined in Table 77. The values are reported according to progression status (progression free or progressed disease) and are based on pooled HRQoL data from the cabozantinib + nivolumab arm and the sunitinib arm of CheckMate 9ER (using the latest data cut provided to the EAG). The EAG noted that utility values for the progression free health state remained relatively high for most subgroups (with the exception of the poor prognostic subgroup) and that for each prognostic subgroup the difference in utility from moving from progression free to progressed disease was relatively minor. Furthermore, utilities for the progression free and progressed disease health states were high relative to those values used in published NICE TAs i.e. 1st line treatments in previously untreated patients (see Section 4.3.7.2).

The company was asked to comment on the face validity of the CheckMate 9ER values relative to those reported within the NICE TAs in Table 24 of the company submission (TA512, TA542, TA581 and TA645). Based on the response provided to the EAG the company were unable to adequately provide a satisfactory explanation, however noted that high utility values were reported in published literature, including Ambavane et al. (2020)[231], Bensimon et al. (2020)[232], McCrea et al. (2018)[164], Haddad et al. (2020)[233] and NICE TA630.[234] The EAG noted these studies to be associated with limitations which prevent the generalisability of values including differences in patient population baseline characteristics, differences in utility estimation methods and lack of robust HRQoL methodology and reporting. Ambavane (2020) report a higher utility than the CheckMate 214 publication despite the authors saying the values are from CheckMate 214, Bensimon (2020) use a time to death approach, McCrea (2018) reports a lack of HRQoL data collection as a limitation of the analysis, Haddad (2020) is in head and neck cancer and TA630 is in NTRK fusion positive tumours.

To further justify the face validity of the CheckMate 9ER utility values study, the company stated that utilities from CheckMate 9ER were supported by the *'rapid and sustained improvement in clinical an HRQoL outcomes'* associated with the mechanism of action of cabozantinib + nivolumab (reference to the MMRM analysis using the median 32.9 month follow up data cut were provided to support this statement). The company also presented time to definitive deterioration data from CheckMate 9ER to support the thesis that cabozantinib + nivolumab reduced the risk of deterioration relative to sunitinib. The EAG noted that whilst cabozantinib + nivolumab resulted in a significant reduction in the risk of deterioration compared to sunitinib using the EQ-5D-3L VAS [HR 0.74 (0.59-0.92)], when the EQ-5D-3L UK utility index was used the difference was non-significant [HR 0.86 (0.70-1.06)]. Additionally, treatment was not selected in the MMRM as a covariate, suggesting that treatment may not meaningfully contribute to the variability of EQ-5D-3L.

The EAG acknowledged the HRQoL data collected and presented in CheckMate 9ER, however, the company's response did not sufficiently postulate why values from the pivotal study were higher than those reported in the majority of other NICE TAs for 1st line treatment of aRCC. Furthermore, based on clinical opinion provided to the EAG, the values from CheckMate 9ER were considered to lack face validity when compared to those reported in other trials including CheckMate 214 and JAVELIN Renal 101. Clinical opinion noted that values from JAVELIN Renal 101 may better reflect patients HRQoL in clinical practice (see Table 78). Additionally, the EAG noted that the utility values estimated from CheckMate 9ER were broadly similar to the

age and sex adjusted EQ-5D-3L values reported by Hernandez Alva et al. (2022), which estimated expected EQ-5D-3L values for UK males and females using the Health Survey England (HSE) 2014 dataset. Baseline utility for males and females aged 61 were estimated to be 0.8476 and 0.8206 respectively, and for males and females aged 62, baseline utility was estimated to be 0.8444 and 0.8165 respectively. Due to the lack of clinical plausibility (and concerns surrounding the MMRM approach), the EAG did not use the company's trial derived utilities in the base case model. However, to test uncertainty, values from CheckMate 9ER have been used in a scenario analysis (see Section 4.3.7.3 for further detail).

Table 77: Utility values from CheckMate 9ER

Risk group	Progression free (mean)	Progressed disease (mean)
ITT	[redacted]	[redacted]
Favourable	[redacted]	[redacted]
Intermediate	[redacted]	[redacted]
Poor	[redacted]	[redacted]
Intermediate and Poor	[redacted]	[redacted]

Abbreviations: ITT, Intention to treat. Utilities were derived from Table 3 in the 'utility and disutility' tab within the company's excel model provided 9th of May 2023. Note: utility values have been marked academic in confidence (AIC) as per the marking within the company submission

4.3.7.2. Literature search and data extraction

A total of 82 studies were identified in the literature containing utility values for people with advanced RCC (1st, 2nd and subsequent lines of therapy). To identify relevant and generalisable utility values for inclusion within the model, a set of prioritisation criteria was established. Based on this criteria, UK and NICE technology appraisals, European and Western (non-European) studies containing utility values (published from 2017 onwards) were considered most relevant for consideration. Using the prioritisation criteria, 34 studies were identified. For the complete list of prioritised studies including rationale for inclusion/exclusion, see the utilities data extraction grid in Appendix D.

- UK studies from 2017 including NICE TAs (n=12)
- Europe (non-UK) studies from 2017 (n=8)
- Western studies from 2017 (non-European) (n=14)

Studies considered for data extraction and inclusion within the decision model were those by Meng et al. (2018)[235], Amdahl et al (2017)[236], Porta et al (2021)[237], Henegan et al. (2022)[238],

Motzer et al (2021)[239], Mouillet et al (2017)[240], Cella et al (2019)[122], Cella et al (2021)[241], Cella et al (2022), Cella et al. (2022)[242], Bedke (2022)[243], Buckley (2019).[244] A summary of results can be found Appendix H). However, these studies were ultimately excluded from consideration due to values not being reported in a manner suitable for model input, the lack of face validity, use of secondary data sources for utility estimates, no direct elicitation from patients and lack of EQ-5D-5L mapping.

Ten published NICE TA's were identified that met the prioritisation criteria (Table 78. The EAG noted that some utility data were not available in the public domain as these were marked as confidential. There was some variability in progression free and progressed utilities across NICE TAs for 1st line treatments (and amongst 2nd line treatments), this appeared to be due to heterogeneity across clinical trials with respect to patient characteristics including risk score. Utilities within these appraisals were presented primarily according to health state/progression status, however in TA650 a time to death (TTD) approach was used. Treatment specific utility values were not commonly used within NICE aRCC appraisals, though this approach was adopted in TA780. In order to be congruent with aRCC TAs submitted to NICE, our model estimates utility based on health state/progression status. Furthermore, NICE TAs were considered as the primary source for utility data for 1st and 2nd line treatments, specifically TA645 and TA498 respectively (see Section 4.3.7.3 for more detail).

Table 78: Utility values in published NICE TAs

TA	Year	Recommendation Population	Intervention	Source of utilities	Utilities
TA858	2023	1L	Pem+lenv	CLEAR trial (EQ-5D-3L)	Redacted
TA830	2022	Adjuvant: increased risk of recurrence after nephrectomy	Pem	KEYNOTE 564 (EQ-5D-5L mapped to EQ-5D-3L)	Disease free: 0.868 PFS (distant metastases): 0.803 PD (distant metastases): 0.772
TA780 (CDF review of TA581)	2022	1L int/poor risk	Nivo+ipi	CheckMate 214 (EQ-5D-3L)	PFS on/off nivo+ipi: 0.793 on and 0.749 off PFS on/off suni: 0.754 on and 0.707 off PPS on/off nivo+ipi: 0.794 on and 0.702 off PPS on/off suni: 0.763 on and 0.707
TA650	2020	1L	Pem+axi	Manufacturer derived utility values from KEYNOTE 426 (EQ-5D-3L). A time to death approach was used in the company's base case.	Redacted NICE noted that use of utilities from KEYNOTE 426 and published literature were acceptable for decision making.
TA645	2020	1L	Ave+axi	JAVELIN Renal 101 (EQ-5D-5L mapped to EQ-5D-3L)	PFS: 0.753 PD: 0.683
TA542	2018	1L int/poor risk	Cabo	TIVO-1(EQ-5D-3L)	PFS: 0.726 PD: 0.649
TA512	2018	1L	Tivo	TIVO-1 (EQ-5D-3L)	PFS: 0.726 PD: 0.649
TA498	2018	2L (1 prior VEGF, ECOG PS 0-1)	Lenv+evero	AXIS (EQ-5D, version unclear)	PFS: 0.69 PD: 0.61
TA463	2017	2L/3L (Prior VEGF)	Cabo	METEOR (EQ-5D-5L)	PFS: 0.817 PD: 0.777
TA432	2017	2L	Evero	Swinburn et al (2010)[245]	SD: 0.795 PD: 0.36

Abbreviations: 1L, 1st line; 2L, 2nd line; 3L, 3rd line; CDF, Cancer Drugs Fund; ECOG, Eastern Cooperative Oncology Group Performance Status; EQ-5D-3L, EuroQol five dimension three level; EQ-5D-5L, EuroQol five dimension five level; NICE, National Institute for Health and Care Excellence; PFS, progression free survival; PD, progressed disease; SD, stable disease; TA, technology appraisal; VEGF, vascular endothelial growth factor

4.3.7.3. Utilities used in the model

As noted previously, the most appropriate sources identified for the base case analyses were TA645 for patients treated at 1st line and TA498 for patients treated at 2nd line. We opted to derive utilities from these NICE TAs on the basis that the utilities for 1st and 2nd line demonstrated face validity, were elicited directly from patients using the EQ-5D and were previously assessed and accepted by NICE. In TA645, quality of life data were collected directly from patients in the JAVELIN Renal 101 study using the EQ-5D-5L. Values were then appropriately mapped to the EQ-5D-3L using the Van Hout crosswalk algorithm,[246] resulting in a PFS utility of 0.753 and a PD value of 0.683. These utilities are in broad alignment with the utilities used in TA512 for tivozanib, the off-treatment values in TA780 for nivolumab + ipilimumab (which derived values from CheckMate 214) and TA542 for cabozantinib. Utilities also reflect clinical opinion to the EAG (which noted that JAVELIN Renal 101 appeared to better reflect patient HRQoL in clinical practice). We noted that in TA498, utilities were not collected in the pivotal trial HOPE 205 and that the values used within that appraisal were taken from the AXIS trial (for axitinib), however the EAG and NICE concluded that utilities from AXIS were appropriate for use in the analysis. We noted that PFS utility in TA498 for 2nd line treatment (0.69) was slightly higher than the PD utility reported in TA645 for 1st line treatment (0.683), thus presenting a logical inconsistency. To mitigate this, our analysis therefore assumes that progression free patients at 2nd line will have a utility of 0.683, reflective of progressed 1st line patients.

To estimate the PD utility in 2nd line and subsequent lines, we used the approach outlined in NICE DSU12 guidance,[247] which states that when utility values from cohorts with combined health states are not available*, 'the multiplicative method should be used to combine the data from subgroups with the single health conditions (p.22)'*. In our analysis, the % reduction in utility (from moving from PFS to PD) in TA498 was used applied i.e. 2nd line utility was estimated as follows 0.69/0.683*0.61=0.616. Due to a lack of robust, published utility values for people receiving 3rd line treatment (or later), the same approach was used to estimate PD utility in later lines. Overall, the decision to apply the percentage reduction in utility (in moving from PFS to progressed disease) from TA498 to estimate utility values for progressed disease at 2nd, 3rd and 4th line, was to ensure logical consistency based upon clinical feedback, that is, to ensure patient utility decreases with disease progression.

For 3rd line, the PFS utility value was assumed to be reflective of the progressed disease value for 2nd line patients, that is 0.616. As described previously, to estimate the progressed disease value, we applied the percentage reduction in moving from PFS to progressed in TA498, to the PFS utility value, which resulted in a 3rd line progressed disease utility value of 0.545. For 4th line, the PFS utility value was assumed to be reflective of the progressed disease value for 3rd line patients, that is 0.545. To estimate the progressed disease value we applied the percentage reduction in moving from PFS to progressed disease in TA498, to the PFS utility value, which resulted in a 4th line progressed disease utility value of 0.482. This value is consistent with palliative care utility estimates within oncology submissions to NICE.

For completeness, the EAG sought clinical input on the validity of this approach. Based on clinician input, the application of a similar proportional decrease in quality of life for each later line of treatment (to that between PFS and PD in 2nd line) may be considered somewhat conservative, as there is likely to be a higher proportional decrease on progression after each line of therapy. In order to explore uncertainty surrounding utility values in later lines (3rd and 4th line), the EAG has conducted scenario analysis assuming a higher proportional decrease in quality of life (see below).

Table 79: Utility values used in the model

Line of treatment	Utility	Source
1L	PFS: 0.753 PD: 0.683	JAVELIN Renal 101(TA645[46])
2L	PFS: 0.683 PD: 0.616	PFS utility assumed to reflect PD in 1L. PD value estimated based on % reduction from the AXIS trial (TA498[56])
3L	PFS: 0.616 PD: 0.545	Estimated based on % reduction from the AXIS trial (TA498).Approach follows NICE DSU12 guidance[247])
4L	PFS: 0.545 PD:0.482	Estimated based on % reduction from the AXIS trial (TA498).Approach follows NICE DSU12 guidance[247])

Abbreviations: PFS, Progression free survival; PD, Progressed disease

Due to a lack of published HRQoL data for carers and to be consistent with previous NICE appraisals for advanced RCC, our analysis did not include carer disutility.

Utility values were adjusted for age and sex using the published equation by Ara and Brazier et al (2010)[248] and the Health Survey England (HSE) 2014 dataset, as per Hernandez Alava et al (2022).[249]

Disutility associated with adverse events has been included in the EAG's model. These were derived from HRQoL data collected in the CheckMate 9ER study (received by the EAG on the 9th of May 2023). Adverse events were included as a variable in the company's MMRM model, which was used to estimate the disutility associated with any grade 3-4 adverse event. The mean disutilities associated with Grade 3-4 adverse events are outlined in Table 80. The EAG noted that several adverse events had a positive impact on patient utility which lacked face validity i.e. neutropenia and hypophosphatemia. Data were not available for specific adverse events within TA858 and given the results of the analysis of CheckMate 9ER these events were expected to be of limited impact, therefore we did not include these adverse events in the model.

The EAG noted that several specific adverse events resulted in relatively high disutility, including anaemia, palmar-plantar erythrodysesthesia (hand/foot) syndrome and fatigue. Based on clinical expert opinion to the EAG, treatment related toxicities accumulate over time, particularly fatigue. Patients can experience fatigue either on an immunotherapy (IO) or TKI, however TKI toxicities are chronic and will impact most patients. For completeness, the EAG has conducted two scenario analyses surrounding adverse event disutilities (see Section 4.3.7.4)

The impact for of the 3 key adverse events was presented to Dr Larkin to check its validity. He stated that the information presented showed impact in the wrong ordering which is likely due to sicker patients being unable to complete the relevant questionnaires. He considered that in fact diarrhoea has the greatest impact, followed by HFS and then fatigue. Given this the utility values for fatigue and diarrhoea from CheckMate 9ER were switched around.

Table 80: Modelled disutility associated with adverse events from CheckMate 9ER

	Disutility (Mean)	**Duration (days)**	**Source**
General Grade 3-4 adverse event disutility	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Specific adverse event (Grade 3-4)			
ALT increased	[redacted]	[redacted]	Assumed same as increased lipase (in line with TA858[38])
Anaemia	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
AST increased	[redacted]	[redacted]	Assumed same as increased lipase (in line with TA858[38])
Decreased appetite	[redacted]	[redacted]	Assumed same as fatigue

	Disutility (Mean)	Duration (days)	Source
Diarrhoea	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9 input for fatigue used based on expert advice. IO-induced diarrhoea was assumed to last longer based on clinical expert advice
Fatigue	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9 input for diarrhoea used based on expert advice
Hypertension	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Hypertriglyceridemia	[redacted]	[redacted]	Assumed same as increased lipase (in line with TA858[38])
Hyponatraemia	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Hypophosphatemia	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Lipase increased	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Leukopenia	[redacted]	[redacted]	Assumed same as platelet count decreased (in line with TA858[38])
Lymphopenia	[redacted]	[redacted]	Assumed same as platelet count decreased (in line with TA858[38])
Nausea	[redacted]	[redacted]	Assumed same as fatigue
Neutropenia	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Palmar-plantar erythrodysesthesia syndrome	[redacted]	[redacted]	CheckMate 9ER[59]; clarification response document A9
Platelet count decreased	[redacted]	[redacted]	Assumed same as neutrophil count decreased from CheckMate 9ER[59]; clarification response document A9
Proteinuria	[redacted]	[redacted]	Assumed same as increased lipase (in line with TA858[38])
Stomatitis	[redacted]	[redacted]	Assumed same as fatigue
Vomiting	[redacted]	[redacted]	Assumed same as diarrhoea
Weight loss	[redacted]	[redacted]	Assumed same as fatigue
Grade 1-2			
Diarrhoea	[redacted]	[redacted]	Assumed to have 50% of the impact as at Grade 3-4 based on clinical expert advice
Fatigue	[redacted]	[redacted]	
Palmar-plantar erythrodysesthesia syndrome	[redacted]	[redacted]	

*No disutility (i.e., zero disutility) considered in the EAG model

4.3.7.4. Scenario analyses conducted

Due to uncertainty surrounding health state utilities (particularly for later treatment lines), the EAG plan to conduct the following scenario analyses;

- 1st line: Use utility values from CheckMate 9ER. These values reflect direct trial data.
- All lines: Use CheckMate 9ER utility values for all lines i.e. CheckMate 9ER data used for 1st and 2nd line utility values (and no decrement is applied for 3rd and 4th lines).
- 2nd line onwards: Assume the same PFS and PD utility for 2nd, 3rd and 4th line i.e. PFS utility of 0.68 and PD utility of 0.616. This is a simplifying assumption, however it is useful to see the impact on the ICER when assuming there is no reduction in HRQoL after 2nd line.
- 3rd and 4th line: Assume a higher proportional decrease in HRQoL on progression from 2nd to 3rd line and from 3rd line to 4th line. This is consistent with clinical advice to the EAG. In this scenario, for 3rd line it will be assumed that the decrease in HRQoL associated with moving from PFS to PD will be 10% more than observed in 2nd line. For 4th line, it will be assumed that the decrease in HRQoL associated with moving from PFS to PD will be 20% more than observed in 3rd line.
- Removing the impact of adverse events: Applied to test the impact of adverse events on the ICERs given that there is the potential for some double counting as utility data comes from trials where a proportion of patients will have experienced adverse events.
- Increase adverse event disutilities by 10%. Applied to test the impact of increasing adverse event disutilities on the ICER. Based on clinical input to the EAG patients are likely to experience disutility due to adverse events. This analysis assumes the impact of these disutilities increases by 10%.

4.3.8. Resource use and costs

4.3.8.1. Results from literature search and data extraction

A total of 13 studies were identified in the literature containing cost and resource use data (Section 4.1.1.3, Figure 35) for people with advanced RCC across different lines of therapy (namely 1st, 2nd and subsequent lines), of which there were ten NICE TAs and three published studies. Subsequent data extraction from these studies was performed. All of the identified studies were found to be UK based and adopted an NHS and PSS perspective. The costs included comprised of drug and administration costs, disease management or health state costs based on the healthcare resource utilised and terminal care costs. Some studies also reported adverse event costs and subsequent therapy costs. Resource use frequency was sourced from one of the following sources: clinical trial or its post-hoc analysis, previous NICE technology appraisals or feedback from clinical experts. Unit costs associated with the healthcare resource use were derived from NHS reference costs and Unit costs of Health and Social Care from

PSSRU etc. Summary of cost and resource use information from published studies has been provided in Table 81 and from previous NICE technology appraisals has been provided in Table 82. Detailed data extraction tables are provided in Appendix D.

It can be noted that the source of unit costs, medicine costs and terminal costs were consistent across the published studies as well as the previous NICE technology appraisals. However, the source of resource use frequency was quite varied across the studies. Table 83 in Section 4.3.8.2, therefore compares the different sources for resource use inputs and provides rationale for selecting specific inputs.

Further, in the following sections, the selection of appropriate sources and specific inputs for each type of costs used in the model has also been discussed briefly.

Table 81: Summary of cost and resource use information from published studies

	Amdahl 2017	**Edwards 2018 [NICE TA463]**	**Meng 2018**
Setting/country	UK	UK	England, UK
Intervention	Pazo	For patients who have received previous cytokine therapy (aldesleukin or interferon alfa): axi, sora, suni, BSC For people who have received previous VEGF-targeted therapy: axi, cabo, evero, nivo, suni	Cabo
Comparator	Suni	The interventions listed above compared with each other and BSC	Axi Evero Nivo
Patient population	Treatment-naïve patients with mRCC consistent with that of the COMPARZ trial	Patients with previously treated aRCC who received previous VEGFR-targeted therapy	Adult patients with aRCC following prior VEGFR-targeted therapy
Cohort/Sample size	1,100 (COMPARZ)	Sample size of the included studies ranged from 14 to 362	1,096
Perspective	NHS and PSS	NHS and PSS	NHS and PSS
Price year	2014	2015	2017 (not explicitly stated but assumed, as prices were inflated to 2017)
Currency	GBP	GBP	GBP
Discount rate	3.5%	3.5%	3.5%
Type of costs included	Costs of treatment initiation, medication, and dispensing for pazo and suni Pre-progression follow-up and monitoring, other mRCC-related care associated with pazo and suni treatment during PFS, post-progression supportive care, and in a sensitivity analysis, post-treatment anti-cancer therapy	Drug and administration costs Disease management costs Terminal care costs Adverse events costs and Subsequent therapy costs	Drug and administration costs Disease management/health state costs Terminal care costs and Adverse events costs

	Amdahl 2017	**Edwards 2018 [NICE TA463]**	**Meng 2018**
Source of resource use estimates	HCRU data sourced from post-hoc analysis of COMPARZ trial.[250] Data collected included medical office visits, laboratory visits and tests, home healthcare, hospitalization, urgent care, and medical/surgical procedures.	Previous NICE TAs complemented by expert clinical opinion sought by AG	Source of resource use frequency not reported
Source of unit costs	National Schedule of Reference Costs for 2011–2012,[251] adjusted to 2014 prices using the Consumer Price Index (CPI) for health.[252]	NHS reference costs 2014-15,[253] PSSRU 2015[254]	NHS reference costs 2014-15,[253] PSSRU 2015[254]
Source of medicine costs	List prices of pazo and suni from BNF. For pazo, the list price was adjusted to reflect 12.5% PAS discount[50] and for suni the first treatment cycle (i.e., 28 days of treatment in first 6 weeks) was provided at no cost.[49]	BNF	BNF Dosing and administration schedules from relevant trials, publications, or NICE TAs[58,85,255]
Source of terminal care costs	Terminal care costs not considered	Based on Nuffield Trust report 2014[256]	Based on Nuffield Trust report 2014

Abbreviations: AG, Assessment Group; aRCC, advanced Renal Cell Carcinoma; BNF, British National Formulary; BSC, Best supportive care; GBP, British Pounds; HCRU, Medical Resource Use; NHS, National Health Services; PAS, Patient Access Scheme; PSS, Personal Social Services; PSSRU, Personal Social Services Research Unit; TA, Technology appraisal; UK, United Kingdom; VEGFR, Vascular Endothelial Growth Factor Receptor;

Table 82: Summary of cost and resource use information from previous NICE technology appraisals

NICE TA #	Year	Patient population	Type of costs included	Source of resource use estimates	Source of unit costs	Source of medicine costs	Source of terminal care costs
TA858	2023	1L int/poor risk, where nivo + ipi would otherwise be offered	Drug costs, Admin and health state costs, AE costs, End of life costs	TA650	PSSRU 2020, NHS reference costs 2019-20	BNF	Based on Nuffield Trust report 2014 inflated to 2019/20 costs
TA830	2022	Adjuvant: increased risk of recurrence after nephrectomy	Drug acquisition costs, administration costs, disease management costs, costs for managing adverse events, subsequent treatment costs and terminal care costs incurred at the end of life	KEYNOTE 564, TA650, clinical expert opinion	PSSRU 2020, NHS reference costs 2019-20	BNF, Dosing from SmPC	Based on Nuffield Trust report 2014 inflated to 2019/20 costs
TA780	2022	1L int/poor risk	Drug costs, Admin and health state costs, AE costs, End of life costs	TA581	Not reported	BNF	Not reported
TA650	2020	1L (not recommended)	Drug acquisition and administration of 1L and subsequent treatments, with adjustment for dose intensity; monitoring and disease management in PF and PD states; treatment of included TEAEs for 1L treatments; and terminal care costs in the last cycle before death	TA542 and clinical expert opinion	PSSRU 2018 and NHS reference costs 2017-18	BNF, dosing from SmPC	Based on Nuffield Trust report 2014 inflated to 2019/20 costs
TA645	2020	1L	Drug costs, Admin and health state costs, AE costs, End of life costs	Aligned with TA581	PSSRU 2018, NHS reference costs 2017-18	BNF	Addicott et al. 2008
TA581	2019	1L int/poor risk	Drug and admin costs, health state costs, subsequent treatment costs and AE costs	TA333 and TA417	PSSRU 2015 and 2017, NHS reference costs 2015-16 and 2016-17	BNF	Based on Nuffield Trust report 2014, inflated to 2016/2017
TA542	2018	1L int/poor risk	Drug and treatment costs, health state unit costs and resource use, AE costs and resource use, Subsequent	Estimated by UK clinicians, aligned with	PSSRU 2016, NHS reference costs 2016-17	BNF	Based on Nuffield Trust report 2014, inflated to 2017

NICE TA #	Year	Patient population	Type of costs included	Source of resource use estimates	Source of unit costs	Source of medicine costs	Source of terminal care costs
			treatment costs and Terminal care costs	TA512 and TA215			
TA512	2018	1L	Drug and treatment costs, health state unit costs and resource use, AE costs and resource use, Subsequent treatment costs	TA333	PSSRU 2015, NHS reference costs 2015-16	BNF	Not reported
TA498	2018	1 prior VEGF, ECOG 0-1	Drug and treatment costs, health state unit costs and resource use, AE costs and resource use, Subsequent treatment costs and Terminal care costs	TA333	PSSRU 2015, NHS reference costs 2015-16	BNF	Based on Nuffield Trust report 2014, inflated to 2016
TA463	2017	Prior VEGF	Drug and treatment costs, health state unit costs and resource use, AE costs and resource use, Subsequent treatment costs and Terminal care costs	Estimated by UK clinicians	PSSRU 2015, NHS reference costs 2015-16	BNF	Based on Nuffield Trust report 2014, inflated to 2016
TA432	2017	Prior VEGF	Drug and treatment costs, health state unit costs and resource use, AE costs and Terminal care costs	SLR and economic evaluation, 2008[257]	PSSRU 2015, NHS reference costs 2014-15	BNF	Guest et al. and Coyle et al.

Abbreviations: AE, Adverse events; BNF, British National Formulary; NHS, National Health Services; PSSRU, Personal Social Services Research Unit, SmPC, Summary of Product Characteristics; TA, Technology appraisal.

4.3.8.2. Disease management or health state costs

The quantum of health state resource use (i.e., medical oncologist outpatient consultations, CT scans, blood tests etc.) was found to differ across the included studies. A comparison especially of the consultant outpatient follow-up and CT scans pre- and post-progression between the estimates from previous NICE TAs[38,52,55] which had detailed description of the health care resource use with the individual components broken down and the BMJ and ESMO published RCC guidelines,[42,43] has been presented below in Table 83. As can be seen, a noticeable variation was observed in the resource use frequency within the NICE TAs and when compared to the published guidelines as well. For instance, while the ESMO RCC guideline recommended a consultant follow up visit every 2-4 months, BMJ RCC guideline indicated that it could be best judged by the treating clinician and in the previous NICE TAs the observed frequency of follow up visit ranged from every month to every three months.

Table 83: Comparison of long term follow up frequency across key published studies/NICE TAs and RCC guidelines

Heath state	Resource type	Resource use frequency				
		NICE TA463[55]	**NICE TA542[52] & TA858[38]**	**Edwards 2018[258]**	**BMJ RCC guideline[42]**	**ESMO RCC guideline[43]**
Pre-progression (on and off treatment	Consultant outpatient follow up	0.67 per 4-week cycle (~every 6 weeks)	0.25 per week (~every month)	Every 3 months	Left to judgement of treating clinician	Every 2 to 4 months
	CT scan	0.33 per 4-week cycle (~every 3 months)	0.08 per week (~every 3 months)	Every 3 months	Few monthly intervals	Every 2 to 4 months
Post-progression (off treatment)	Consultant outpatient follow up	Not included*	0.25 per week (~every month)	Not included	Left to judgement of treating clinician	Every 2 to 4 months
	CT scan	Not included*	0.08 per week (~every 3 months)	Not included	Few monthly intervals	Every 2 to 4 months
	GP and specialist nurse visit	1 per 4-week cycle (every month)	Not applicable	20 visits per year (only specialist nurse visit)	Not discussed	Not discussed

Abbreviations: BMJ, British Medical Journal; ESMO, European Society for Medical Oncology; NICE, National Institute of Health and Care Excellence; RCC, Renal Cell Carcinoma, TA, Technology appraisal;

*TA463 was conducted in previously-treated patients at a time where few options were available, therefore post-progression here essentially represents BSC and patients were assumed to be discharged from the oncology.

Note: There was no clear reason reported for why there is a difference in resource use frequency between NICE TA463 and Edwards 2018 (the related EAG monograph), however, it looks likely that the clinical expert opinion to EAG matured over time as Edwards 2018 indicated that estimates based on TA333 and TA417 were complemented by clinical expert opinion to AG (however such a statement was not explicitly available in NICE TA463)

The health state costs and resource use estimates used in the model (Table 84) were based on NICE TA542[52], TA858[38] and Edwards 2018,[258] also complemented by the clinical expert opinion to EAG.

When initiating a new line of treatment patients would have an initial visit with the medical oncologist (including a blood test) and a specialist nurse visit happening alongside. Then a subsequent visit where tolerability to the new treatment would also be assessed (in line with standard practice of a formal medical review to determine tolerability[37]), followed by successive follow up visits. It is to be noted that given the advanced stage of the disease and acknowledging some patients might need to be seen more or less frequently, a monthly follow up until 12 weeks and every 2.5 months beyond 12 weeks based on clinical opinion to EAG was deemed appropriate.

Patients would also receive CT scans every 3 months (which was found to be almost consistent across the included studies) to check for the signs of progression and a routine blood test aligned with the consultant visits. The frequency of consultant follow-up visits, CT scans and blood tests was assumed to be the same across all lines of treatment, as monitoring would broadly remain the same irrespective of the treatment received (consistent with NICE TA858[38]). In addition, patients were assumed to have daily pain medication and regular specialist nurse visits in line with Edwards 2018,[258] however, only during the last line of treatment prior to death. These assumptions were also checked with the clinical experts.

Table 84: Health states resource use and unit costs

Health state	Resource type	Frequency of use (per week)	Unit cost (2022 costs)	Source
Treatment initiation	Consultant outpatient visit (first visit)	1	£206.47	Frequency: NICE TA858 Unit cost: NHS reference costs 2021-22; HRG code WF01B, Clinical oncology - Non-Admitted Face-to-Face Attendance, First
	Specialist nurse visit	1	£53	Frequency: assumed same as consultant visit per clinical opinion to EAG Unit cost: PSSRU 2022,[259] Section 11.2.2 Nurse specialist (Band 6), cost per working hour
	Blood test	1	£2.39	Frequency: NICE TA 858 Unit cost: NHS reference costs 2021-22; HRG code DAPS 03 - Integrated blood services
All lines of treatment, on and off treatment (until 12 weeks)	Consultant outpatient follow up	0.25 (until 12 weeks) 0.1 (beyond 12 weeks)	£164.19	Frequency: NICE TA542, NICE TA858 until 12 weeks; every 2.5 months beyond 12 weeks based on clinical opinion to EAG Unit cost: NHS reference costs 2021-22; HRG code WF01A, Clinical oncology - Non-Admitted Face-to-Face Attendance, Follow up
	CT scan	0.083	£99.88	Frequency: NICE TA542, NICE TA858 Unit cost: NHS reference costs 2021-22; HRG code Outpatient - RD27Z – CT scan of more than three areas
	Specialist nurse visit	0.25	£53	Frequency: assumed to happen in conjunction with consultant visit per clinical opinion to EAG Unit cost: PSSRU 2022,[259] Section 11.2.2 Nurse specialist (Band 6), cost per working hour
	Blood test	0.25	£2.39	Frequency: NICE TA542, NICE TA858 Unit cost: NHS ref costs 2021-22 DAPS03 Integrated blood services
BSC	Consultant outpatient follow up	0.25	£164.19	Frequency: assumed to happen in conjunction with specialist nurse visit based on clinical opinion to EAG Unit cost: NHS reference costs 2021-22; HRG code WF01A, Clinical oncology - Non-Admitted Face-to-Face Attendance, Follow up
	Specialist nurse visit	0.25	£53	Frequency: Based on Edwards 2018 but assumed to be twice as frequent as consultant follow up

Health state	Resource type	Frequency of use (per week)	Unit cost (2022 costs)	Source
				Unit cost: PSSRU 2022,[259] Section 11.2.2 Nurse specialist (Band 6), cost per working hour
	Pain medication	7 (1 mg/ml vial morphine sulphate daily)	£5.78	Frequency: Based on Edwards 2018 Unit cost: BNF; 50mg/50ml vial morphine sulphate solution for infusion

Abbreviations: BNF, British National Formulary; NICE, National Institute of Health and Care Excellence; NHS, National Health Services; PSSRU, Personal Social Services Research Unit; TA, Technology appraisal;

4.3.8.3. End of life costs

End of life or terminal care costs are incurred by all patients dying in the model based on the Nuffield Trust report exploring the cost of care at the end of life.[256] All the previous published studies and the NICE TAs (except TA645) derived terminal care cost from this report (as seen in Table 82).

The cost components of terminal care per the Nuffield Trust report have been given below in Table 85. All costs are presented from an NHS / PSS perspective and were inflated to 2022 costs using the NHS cost inflation indices (NHSCII) from PSSRU.[259] The total estimated cost of terminal care (inflated to 2022) was found to be £8,714.

Table 85: Summary of costs related to end of life or terminal care

Resource type	Resource use frequency*, Mean (SD)	Unit cost per patient (SD, where available)	Source	Total costs (adjusted for inflation)
GP consultation	11.4 (6.2) visits	£42	Resource use frequency: Nuffield Trust report, 2014.[256] [Table 1, Group: Cancer diagnosis] Unit cost: PSSRU 2022,[259] Section 9.4 GP unit costs – patient contact lasting 9.22 minutes, including direct care staff and with qualification costs	£479
District nursing care	7.53 (19.57) hours	£53	Resource use frequency: Nuffield Trust report, 2014. [Table 2, Group: Cancer diagnosis] Unit cost: PSSRU 2022,[259] Section 11.2.2 Nurse specialist (Band 6), cost per working hour	£399
Local authority funded social care	Not available	£444 (£1,484)	Cost: Nuffield Trust report, 2014. [Table 3, Group: Cancer diagnosis; 2010 costs]	£549
Hospital care	Not available	£5,890 (£5,264)	Cost: Nuffield Trust report, 2014. [Table 4, Group: Cancer diagnosis; 2010 costs]	£7,287
Total				£8,714

Abbreviations: GP, General Practitioner; PSSRU, Personal Social Services Research Unit; SD, standard deviation

* number of visits or cost of care in the last 90 days before death

Note: 2010 costs were inflated to 2022 by applying year on year annual % increase on the 2014/15 HCHS index = 293.1 from PSSRU 2017[260] (which resulted in 2022 index = 332.3)

4.3.8.4. Drug and administration costs

A summary of acquisition costs of the treatments considered in the 1st line setting and their respective dosing schedules (as provided in detail in Section 4.3.3), along with the treatments in subsequent lines has been presented in Table 86 below. Please note that the unit costs for each drug were extracted from either the drugs and pharmaceutical electronic market information tool (eMIT) or the British National Formulary (BNF) and the cheapest unit price was used where multiple formulations existed for the same drug. Except for everolimus and sunitinib (for which the costs were derived from eMIT), all other drug costs were sourced from BNF.

The per cycle costs for each drug component were calculated based on the respective dosing regimen/intensities and were applied to proportion of patients remaining on treatment in each model cycle within the modelled time horizon (informed by the TTD curve in the base case and mean number of administrations in the scenario analysis). The dosing regimens are the same across the favourable and intermediate/poor risk subgroups and RDIs are assumed equivalent across subgroups.

Wastage is calculated for IV administered drugs dosed by patient weight with the average number of vials calculated using the method of moments based upon the subset of patients for whom individual patient weights were available within the UK RWE (patients who received nivolumab + ipilimumab). The model base case considers wastage with the assumption of no wastage explored in scenario analysis. Considering wastage increased the cost of nivolumab by 4% and the cost of ipilimumab by 30%. Further, for IV drugs given at a fixed dose missed doses were assumed not to be wasted in the base case based upon expert clinical input that steps are taken to minimise wastage and that either the shelf life is so short that treatments are only prepared when a patient has confirmed attendance (ipilimumab) or remaining vials are reused (other products). For oral treatments, no additional wastage costs were included as costing was done based on packs used.

The model will include confidential PAS and commercial access arrangement discounts (where applicable) as received from NICE with the ICER containing all discounted prices presented in a confidential appendix.

Table 86: Acquisition costs of treatments considered in the economic model

Treatment	Formulation	Size of pack	Dose per unit	Pack price (list price)[261,262]	Confidential discount price (discount %)
Ave	Bavencio® 200 mg/10 ml infusion vials	1 vial	20 mg per ml	£768	See cPAS appendix
Axi	Inlyta® 5 mg tablets	56 tablets	5 mg	£3,517	
Cabo	Cabometyx® 40 mg	30 tablets	20, 40 and 60 mg	£5,143	[redacted]
Evero	Evero 10 mg tablets (generic)	30 tablets	10 mg	£373.48	See cPAS appendix
Ipi	Yervoy® 50mg/10 ml infusion vials	1 vial	5 mg per ml	£3,750	
Lenv	Lenvima® 10 mg capsules	30 capsules	10 mg	£1,437	
Nivo	Opdivo® 100mg/10 ml infusion vials	1 vial	10mg per ml	£1,097	
	Opdivo® 40mg/4 ml infusion vials	1 vial	10 mg per ml	£439	
Pazo	Votrient® 400 mg tablets	30 tablets	400 mg	£1,121	
Pem	Keytruda® 100mg/4 ml infusion vials	1 vial	25 mg per ml	£2,630	
Suni	Suni 50 mg capsules (generic)	28 capsules	50 mg	£1,388.77	
Tivo	Fotivda® 1340 µg capsules	21 capsules	1.34 mg	£2,052	

Abbreviations: mg, milligrams; ml, millilitres; NHS, National Health Service; UK, United Kingdom.

Relative dose intensities from trials and RWE (with RWE considered in base case and trial estimates in scenario) are applied to calculate the actual cost of the treatments consistent with the previous NICE technology appraisals, as provided in Table 87. RWE data was not available for cabozantinib + nivolumab, pembrolizumab + lenvatinib or the IO component within combination therapies; in the scenario using RWE we assume these are the same as the trial information available.

Table 87: Relative dose intensities of treatments considered (trial and RWE)

Drug	Treatment line	Relative dose intensity, % (SE where available)		Trial source/assumption
		Trial	RWE	
Ave+axi	1L advanced	Ave: 91.5 Axi: 89.4	[redacted]	Motzer et al 2019[95]

Drug	Treatment line	Relative dose intensity, % (SE where available)		Trial source/assumption
		Trial	RWE	
Axi	Prior TKI or cytokine (2L)	99	[redacted]	AXIS trial: Rini et al. 2011[85]
Axi	3L	99	[redacted]	Assumed same as 2L
Axi	4L		[redacted]	
Cabo	1L advanced	93.3	[redacted]	CABOSUN Clinical study report (as reported in TA542[52])
Cabo	2L	93.3	[redacted]	Assumed same as 1L
Cabo	3L+	93.3	[redacted]	Assumed same as 1L
Evero	Prior VEGF (2L)	84 (1.1)	[redacted]	METEOR clinical study report (as reported in TA542[52])
Evero	3L	84 (1.1)	[redacted]	Assumed same as 2L
Evero	4L		[redacted]	
Lenv+ evero	Prior VEGF (2L)	Lenv: 70.4 Evero: 89.3	[redacted]	CLEAR trial: Motzer et al 2021[45]
Lenv+ evero	3L	Lenv: 70.4 Evero: 89.3	[redacted]	Assumed same as 2L
Lenv+ pem	1L advanced	Lenv: 69.6 Pem: 62.9 – median number of infusions reported as 22	[redacted]	CLEAR trial: Motzer et al 2021[45]
Nivo	Previously treated (2L)	97.5	[redacted]	CheckMate 025 company submission (as reported in NICE TA463[55])
Nivo	3L	97.5	[redacted]	Assumed same as 2L
Nivo+ cabo	1L advanced	Nivo: [redacted] Cabo: [redacted]	[redacted]	CheckMate 9ER (clarification response; A10a)
Nivo+ ipi	1L advanced	Nivo induction: 79*; Nivo maintenance: [redacted] Ipi: 79*	[redacted]	Motzer et al 2018[90] For nivo, same RDI as cabo+nivo to be assumed for nivo mono maintenance as data not available
Pazo	1L advanced	86	[redacted]	VEG105192 trial (as reported in NICE TA215[50] and TA512[51])
Pazo	2L	86	[redacted]	Assumed same as 1L
Pazo	3L		[redacted]	
Suni	1L	[redacted]	[redacted]	CheckMate 9ER (clarification response; A10a)
Suni	2L	[redacted]	[redacted]	Assumed same as 1L
Suni	3L		[redacted]	

Drug	Treatment line	Relative dose intensity, % (SE where available)		Trial source/assumption
		Trial	RWE	
Tivo	1L advanced	94	[redacted]	TIVO-1 study (as reported in NICE TA512[51])
Tivo	2L	94	[redacted]	Assumed same as 1L
Tivo	3L		[redacted]	

Abbreviations: 1L, 1st line; 2L, 2nd line; 2L+, 2nd line-plus; NICE, National Institute for Health and Care Excellence; NR, not reported; RDI, relative dose intensity; SE, standard error; TA, technology appraisal; TKI, tyrosine kinas inhibitor

*79% reported to receive all 4 doses of nivolumab and ipilimumab within the induction phase

Different administration modes were used for different drugs depending on route of administration and whether or not the drug is administered jointly based on NICE TA858/TA645, which has been provided below in Table 88, along with the unit costs extracted from NHS reference costs 2021-22.[259]

Table 88: Unit cost of drug administration

Treatments	Administration mode	Unit cost (2022)	Source
Pem, nivo, ave	Simple parenteral Chemotherapy at First Attendance - Outpatient	£207.59	NHS reference costs 2021-22; HRG code: SB12Z
Ipi (for first 4 cycles when nivo is delivered jointly with ipi)	Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance - Outpatient	£440.71	NHS reference costs 2021-22; HRG code: SB14Z
Lenv, suni, pazo, tivo, axi and cabo	Exclusively Oral Chemotherapy (first cycle) + Pharmacist (Band 6) assuming 12 minutes (subsequent cycles)	First cycle: £197.25 + Subsequent cycles: £11	First Cycle: NHS reference costs 2021-22; HRG code: SB11Z – Deliver exclusively oral chemotherapy. Subsequent cycles: PSSRU 2022. Pharmacist time based on NICE TA645.

Abbreviations: HRG, Healthcare resource group; IV, intravenous; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; TA, technology appraisal
Note: 2020-21 costs were inflated to 2022 using NHSCII annual % increase on previous year index (2.72%) from PSSRU 2022[259]

4.3.8.5. Adverse event costs

AE management costs have been calculated using the unit costs per event and the rate of AEs for each treatment under consideration (for the two options explained in Section 4.3.6).

Table 89 presents the costs per event of all the adverse events considered per the two options/data sources mentioned in Section 4.3.6, incorporating the clinical opinion to EAG, in line with NICE TA858[38] and the unit costs derived from NHS reference costs 2021-22[263].

Table 90 presents the average cost and QALY decrement of Grade 3+ and specified grade 1/2 AEs for each treatment considered in the base case based on RWE. Please note that the similar table for the trial scenario has been presented along with the AE rates from trials in Appendix O. The disutilities associated with the AEs considered have been provided and described in Section 4.3.7.3.

Table 90 was presented to Dr Larkin for comment. He noted that he would have expected tivozanib and axitinib to be more similar given their similar mechanism of action. The ordering of the TKI monotherapies was as expected. Given this a scenario analysis has been included setting the impact of axitinib on adverse events to the same as tivozanib. Dr Larkin also noted that he would have expected similar treatments to be more closely grouped together particularly TKI monotherapy and lenvatinib + everolimus with more AEs than monotherapy and the IO+TKIs with nivolumab monotherapy and nivolumab + ipilimumab expected to be different. This does appear to be the case when looking at the total cost of managing AEs but and QALY impact, but this sensible grouping is not seen when looking at per cycle impacts which validates the choice to use one-off cost and QALY impacts in the base case.

Noting previous clinical advice that the impact of AEs has often been underestimated in previous appraisals, scenario analysis is also presented doubling this impact.

.

Table 89: Adverse event costs per event

AEs	Cost per event (2022 costs)[263]	Assumptions
Grade 3+		
Anaemia	£655.75	Weighted average SA04G-L. Iron Deficiency Anaemia, Non-elective stay + nurse (GP practice) cost per hour
Decreased appetite	£0.00	Assumption
Diarrhoea (TKI induced)	£827.18	Based on clinical opinion to EAG: Weighted average FD10E-H Non-Malignant Gastrointestinal Tract Disorders without Interventions, non-elective short-stay
Diarrhoea (IO induced)	£4,321.12	Based on clinical opinion to EAG: Weighted average FD10E-H Non-Malignant Gastrointestinal Tract Disorders without Interventions, non-elective long-stay
Fatigue	£662.61	Based on clinical opinion to EAG: 3*Consultant led medical oncology service: service code 370 (blood test cost not included as it is already included in resource use)
Hypertension	£424.60	EB04Z. Non-elective short stay.
Hypertriglyceridemia	£0.00	Assumed to be zero as regular blood test already considered in health state costs; in line with TA542[52]
Hyponatraemia	£574.71	Weighted average KC05G-N, Fluid or electrolyte disorders, non-elective short stay
Hypophosphatemia	£0.00	Assumed to be zero as regular blood test already considered in health state costs; in line with TA542[52]
Increased ALT	£0.00	Assumed to be zero as regular blood test already considered in health state costs; in line with TA542[52]
Increased AST	£0.00	Assumed to be zero as regular blood test already considered in health state costs; in line with TA542[52]
Increased lipase	£655.75	Assumed to be the same as anaemia (per TA645)[46]
Leukopenia	£0.00	Assumed to be zero as regular blood test already considered in health state costs; in line with TA542[52]
Lymphopenia or lymphocytopenia	£679.97	Weighted average of SA35A-E Agranulocytosis. Non-elective short stay + nurse (GP practice) cost per hour
Nausea/vomiting	£801.11	Non-elective short stay unit cost assumed.
Neutropenia	£655.75	Assumed same cost as anaemia (as per TA645)[46]
HFS or Palmar-plantar syndrome	£621.43	Based on clinical opinion to EAG: 50% of patients are admitted to a general medical word for a short stay, the other 50% see their oncologist (~2 x appointments)
Platelet count decreased	£801.11	Non-elective short stay unit cost assumed (as per TA498)[56]
Proteinuria	£220.87	Consultant led medical oncology service: service code 370 (as per TA542[52])
Stomatitis	£801.11	Assumed same as weight decreased cost
Weight decreased	£801.11	Non-elective short stay unit cost (as per TA645)

AEs	Cost per event (2022 costs)[263]	Assumptions
Grade 1/2		
Diarrhoea	£220.87	Based on clinical opinion to EAG: Outpatient appointment + blood test to rule out infection (blood test cost not included as it is already included in resource use)
Fatigue	£441.74	Based on clinical opinion to EAG: 2*Consultant led medical oncology service: service code 370 + blood test (blood test cost not included as it is already included in resource use)
HFS or palmar-plantar syndrome	£441.74	Based on clinical opinion to EAG: 2*Consultant led medical oncology service: service code 370

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; HFS, Hand-foot syndrome; TA, technology appraisal

Table 90: Total AE (grade 3+ and grade 1/2) costs (base case)

Treatment	AE costs (per cycle)	AE costs (one-off)	QALY decrement (one-off)
Cabo+nivo	£11.89	£1,126.81	-0.003
Nivo+ipi	£9.75	£334.76	-0.029
Nivo	£6.74	£161.13	-0.006
Lenv+pem	£14.36	£1,061.75	-0.027
Ave + axi	£17.91	£1,035.51	-0.028
Pazo	£14.71	£511.61	-0.013
Tivo	£14.50	£407.94	-0.007
Suni	£15.60	£603.81	-0.013
Cabo (1L)	£20.28	£731.83	-0.017
Cabo (2L)	£20.61	£743.65	-0.025
Lenv+evero	£28.53	£942.87	-0.004
Evero	£20.69	£332.88	-0.038
Axi	£58.72	£1,634.23	-0.017

Abbreviations: AE, Adverse events

4.3.8.6. Subsequent treatment costs

Given different pathways are possible following and conditional upon 1st line treatments received in aRCC treatment landscape, relevant subsequent treatment costs need to be considered upon progression and subsequent treatment discontinuation. Within the state transition analysis subsequent treatment costs are applied to patients on treatment per line of therapy dependent upon the sequence being calculated. Within the PartSA analysis subsequent treatments are applied as a one-off cost on progression based on the mean duration of subsequent treatment.

Two relevant data sources were considered for calculating the subsequent treatment costs:

1. Costs based on subsequent treatments as observed in RWE (see Section 3.5) and
2. Costs based on subsequent treatments from CheckMate 9ER or other relevant comparator pivotal trials (Appendix N)

The UK RWE is used for subsequent systemic therapies in the model base case (Table 92) to better reflect clinical practice with the distribution of subsequent treatments observed in the trials will be explored as a scenario analysis. When analysing the UK RWE treatments which are not available via routine commissioning as illustrated in the treatment pathway diagram (Figure 6) were not included. It is to be noted that the subsequent radiotherapy and surgery costs were also considered (as given in Table 91 below) following progression and added as a one-off cost with frequencies based on data from CheckMate 9ER as data was not available from the UK RWE. Pooled rates from both arms were used as the proportion of patients receiving subsequent radiotherapy and subsequent surgery was similar.

The following assumptions were made to inform the subsequent treatment proportions and durations. The same drug and administration costs were used as described in Section 4.3.8.4.

Assumptions common to both RWE and trial:

- The type of subsequent treatment was assumed to be independent of the 1st line risk group and only dependent on the prior treatments received. Analysis of real-world evidence stratifying the contingency table of treatment types at first and second line (excluding types only available for intermediate/poor risk groups at first line, i.e. IO/IO combination) suggested that this was a reasonable assumption, with no evidence of

interaction between risk group and type of second-line treatment conditional on first-line treatment (p=0.88).

- Subsequent treatment proportions were set to zero for nivolumab after an IO had already been used in line with UK clinical practice for all subsequent lines
- Subsequent treatments after pazopanib and sunitinib were assumed to be the same as tivozanib for 3rd line as data was too sparse to estimate separately
- All subsequent treatment proportions were adjusted based on BSC proportions sourced from RWE and CheckMate 9ER (as it was otherwise unavailable in the trial-based scenario).
- Where the final percentages calculated did not sum to 100% either due to rounding errors, patients receiving sequences that did not follow UK practice or data indicating patients received the same treatment twice patients were reallocated equally between all sequences that involved an active 2nd line systemic treatment.
- Where data were not available for the duration of subsequent treatments from one source data from the alternative source was used (for instance where mean treatment duration was not available from trials, mean duration from the RWE was used instead) – this only impacts scenario analysis using the PartSA model

Table 91: Subsequent radiotherapy and surgery costs following progression

	Unit Cost (£)	Number of sessions	Proportion of patients receiving (all risk)*	Source/Assumptions
Subsequent radiotherapy	£255.51	2	[redacted]	Based on clinical input to the EAG the main uses are palliation for painful mets (particularly bone mets), gamma knife for brain mets or SBRT for oligometastatic disease to postpone resistance to therapy with the last two uses being more expensive but rarer. Most patients require 2 treatments for palliative radiotherapy with SBRT requiring 5 treatments. BMJ guidance[264] states that palliative radiotherapy is inexpensive and generally given at a low dose using a linear accelerator and that it takes around 15 minutes. The cost code selected was aligned to this guidance. SC31Z - Deliver a Fraction of Adaptive Radiotherapy on a Megavoltage Machine; NHS reference costs 2021-22; outpatient
Subsequent surgery	£5,393.26	1	[redacted]	Clinical expert advice was that there are a large number of types of surgery possible. The key types which occur following start of systemic treatment are lobectomy or pneumonectomy if single / two sites in lung nodules, fixation for symptoms for bone fracture in bones or for excisions for single site for bone mets (radical approach), stents to optimise kidneys and occasionally resection for brain metastases, In some cases nephrectomy is deferred until after the patient has started systemic therapy and is used in patients who respond well to systemic therapy to gain better disease free survival figures. Average of weighted costs of HRG codes selected as broadly representative: LB06J-M and DA17P-R (based on clinical opinion to EAG); NHS reference costs 2021-22; elective inpatient

***** Denominator is the number of patients receiving a 2nd line treatment

Table 92: Subsequent treatment proportions from the UK RWE

1L treatments	Subsequent 2L treatments, %							
	Axi	**Cabo**	Lenv+evero	**Nivo**	**Pazo**	**Suni**	**Tivo**	**BSC**
Ave+axi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

1L treatments	Subsequent 2L treatments, %							
	Axi	Cabo	Lenv+evero	Nivo	Pazo	Suni	Tivo	BSC
Cabo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Nivo+ipi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cabo+nivo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lenv+pem	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pazo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Suni	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tivo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

2L treatments	Subsequent 3L treatments, %								
	Axi	Cabo	Lenv+evero	Evero	Nivo	Pazo	Suni	Tivo	BSC
Axi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cabo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lenv+evero	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Nivo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pazo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Suni	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tivo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

3L treatments	Subsequent 4L treatments, %		
	Axi	Evero	BSC
Axi	[redacted]	[redacted]	[redacted]
Cabo	[redacted]	[redacted]	[redacted]
Lenv+evero	[redacted]	[redacted]	[redacted]
Nivo	[redacted]	[redacted]	[redacted]
Suni	[redacted]	[redacted]	[redacted]
Tivo*	[redacted]	[redacted]	[redacted]
Pazo*	[redacted]	[redacted]	[redacted]

Abbreviations: Axi, axitinib; Cabo, cabozantinib; Evero, everolimus; Len, Lenvatinib; Nivo, nivolumab; Pazo, pazopanib; RWE, real world evidence; Suni, sunitinib; Tivo, tivozanib; UK, United Kingdom

*Assumed equal to sunitinib as no 4th line treatments were observed after tivozanib in the dataset

Table 93. Subsequent treatment costs (base case using RWE at list price)

Population	1L treatment	Average one-off drug cost weighted by sub txt prop and mean duration of treatments (PartSA scenario only)	Average one-off admin cost weighted by sub txt prop and mean duration of treatments (PartSA scenario only)
All/fav risk	**Cabo+nivo***	£29,506.35	£650.95
	Ave+axi	£34,024.80	£647.70
	Pazo	£51,823.32	£4,310.21
	Tivo	£54,225.21	£5,122.24
	Suni	£51,333.18	£4,409.59
Int/poor risk	**Cabo+nivo***	£29,506.35	£650.95
	Nivo+ipi	£26,619.16	£750.26
	Pem+lenv	£33,784.61	£663.48
	Ave+axi	£34,024.80	£647.70
	Pazo	£51,823.32	£4,310.21
	Tivo	£54,225.21	£5,122.24
	Suni	£51,333.18	£4,409.59
	Cabo	£47,280.42	£5,836.59

Abbreviations: Axi, axitinib; Cabo, cabozantinib; Evero, everolimus; Len, Lenvatinib; Nivo, nivolumab; Pazo, pazopanib; Suni, sunitinib; Tivo, tivozanib; fav, favourable; int, intermediate

*Cabo+nivo subsequent treatment costs were found to be lower as none of the treatment sequences starting with cabo+nivo in 1L, included nivo or cabo in the subsequent lines for which the drug costs and the treatment duration in subsequent lines were relatively higher

4.3.9. Severity

The NICE manual is unclear as to how current practice should be defined in a multi-comparator decision space such as is present here for calculation of the severity modifier. There are three clear options to define current practice in these circumstances:

- Define a common reference treatment to calculate severity modifiers for all other treatments compared to this
- Calculate the severity modifier based upon the market shares of all the comparators
- Calculate severity modifiers separately for pairwise comparisons

Use of pairwise comparisons, whilst being the simplest option, is inconsistent with the principle of fully incremental analysis. Use of market shares would also be inconsistent with the principle of fully incremental analysis. Therefore, in the EAG base case absolute and proportional shortfall are calculated using a common reference treatment for the overall population and each risk subgroup with QALY weightings assigned based upon NICE's severity modifiers (Table 94). The reference treatment to which cabozantinib + nivolumab is compared is the treatment with the largest absolute QALYs which is not ruled out via the rules of dominance / extended dominance within incremental analysis. The EAG consider this to represent current best practice in the absence for formal NICE guidance. Pairwise analysis will be presented in addition.

Table 94: QALY weightings for severity

QALY weight	Proportional QALY shortfall	Absolute QALY shortfall
1	Less than 0.85	Less than 12
x1.2	0.85 to 0.95	12 to 18
x1.7	At least 0.95	At least 18

Abbreviations: QALY, quality-adjusted life-year

The future health lost by people living with RCC was calculated using age and sex data taken from the UK RWE on an individual patient level to preserve correlations. ONS life tables (2018 – 2020)[230] were used to calculate future life expectancy for the general population and the HSE 2014 dataset to calculate future quality of life for the general population.[249] QALYs for the general population were discounted at a rate of 3.5%, consistent with modelled QALYs for RCC treatments.

Modelled discounted QALYs for the reference treatment were then be used to calculate absolute and proportional QALY shortfall amounts and the relevant QALY modifier to apply.

4.3.10. Uncertainty

Base case analyses will be probabilistic as this generates expected outcomes and costs and is in line with the NICE manual.[74] Additional scenario and one-way sensitivity analyses have been conducted where they add value and clarity.

Table 95: List of scenario analyses conducted

Parameter	Base case	Scenario	Justification
Model structure			
Overall structure	State transition 4 lines	PartSA	Most frequently used structure in prior submissions
		State transition 3 lines	Last line at which there is good sample size in the UK RWE
		State transition 2 lines	Matches number of lines available from CheckMate 9ER
Discount rate	3.5%	0%	NICE manual 2022
		6%	NICE manual 2022
Primary data source			
Data source for baseline risk and patient characteristics	UK RWE, state transition model	Trial-based analyses, state transition model	Testing impact of use of trial data rather than RWE for patient characteristics, baseline risk and subs therapy
	UK RWE, state transition model	Trial-based analyses, PartSA	Testing interaction with model structure
Population characteristics			
Data source	UK RWE	CheckMate 9ER	Testing impact of patient characteristics alone
Use means or IPD	IPD	Mean	Testing impact of use of individual patient characteristics preserving correlation between age and sex vs means
Effectiveness			
Baseline risk	UK RWE	CheckMate 9ER	Testing impact of baseline risk
Preferred 1st line NMA	FP NMA	PH NMA	Testing impact of NMA used
Preferred 2nd line NMA	PH NMA	FP NMA	Testing impact of NMA used
Preferred NMA for pem+lenv	PH NMA	FP NMA	Testing impact of NMA used
Method used to adjust crossing curves	Hazards	Survivor function	No guidance available for preferred method
Assume pazo equal to suni	Yes	No	

Parameter	Base case	Scenario	Justification
Assume tivo equal to suni	Yes	No	Testing impact of relaxing equivalency assumptions from prior TAs
Assume evero equal to axi	No	Yes	
Time on treatment data taken from	TTD	PFS	TTD data are sparse, testing use of PFS which is available for all treatments but less accurate
	Relative effectiveness for nivo + ipi from PFS consistent with other treatments	Relative effectiveness for nivo + ipi from simple HR between PFS and TTD from CheckMate 214	TTD Kaplan Meier supplied by the company indicates a different relationship between TTD and PFS than for other treatments
Treatment effectiveness waning	5 years for IO/TKIs, all endpoints, based on hazards	10 years for IO/TKIs, all endpoints, based on hazards	Considerable uncertainty around long-term relative effectiveness
		10 years all IO combinations, all endpoints, based on hazards	
		Between 5 and 20 years all IO/TKIs, all endpoints, based on hazards	
		Between 5 and 20 years all IO combinations, all endpoints, based on hazards	
		No treatment effect waning	
		PartSA: 5 years for IO/TKIs, OS only, based on hazards	
		Between 4 and 6 years for IO/TKIs, OS only, based on absolute survival	
Survival curve selections			
All risk population			
Sunitinib RWE 1L			
PFS	Log-logistic	Weibull	Good statistical and visual fit. Broadly consistent with external data.
TTD	Log-logistic	Weibull	Good statistical and visual fit. Consistent with CheckMate 214 data Consistent with PFS
TTP	Log-logistic	Weibull	Good statistical and visual fit. Consistency with PFS selection.
OS	Exponential	Weibull	Good statistical and visual fit.

Parameter	Base case	Scenario	Justification
			Midrange estimates within plausible curves.
Cabozantinib RWE 2L			
PFS	Log-logistic	Generalised gamma, Weibull	Good statistical and visual fit. Consistent across lines.
TTP	Log-normal	Weibull	Best statistical and clear best visual fit. Consistent across lines.
OS	Log-logistic	Weibull	Good statistical and visual fit. Consistent across lines.
Cabozantinib RWE 3L			
PFS	Log-logistic	Generalised gamma, Weibull	Good statistical and visual fit. Consistent across lines.
TTP	Log-normal	Log-logistic, Generalised gamma	Good statistical and visual fit. Consistent across lines
OS	Log-logistic	Weibull	Good statistical and visual fit. Consistent across lines.
BSC			
4th line PPS pooled	Log-normal	Exponential	All curves provide similar AUC due to completeness of KM data. Consistency with expert elicitation Note Kaplan Meier based on 19 patients
Intermediate/poor risk population			
PFS	Log-logistic	Weibull, Log-normal	Good statistical and visual fit. All feasible curves provide similar AUC Broadly consistent with external data and expert elicitation Consistent with all risk population
TTD	Log-logistic	Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Consistent with PFS Consistent with all risk population
TTP	Log-logistic	Weibull	Good statistical and visual fit. All feasible curves provide similar AUC. Consistency with PFS selection. Consistent with all risk population.
OS	Exponential	Weibull	Good statistical and visual fit. Consistent with all risk population.
Favourable risk population			

Parameter	Base case	Scenario	Justification
PFS	Log-logistic	Weibull	Good statistical and visual fit. All feasible curves provide similar AUC Broadly consistent with external data and expert elicitation Consistent with all risk population
TTD	Log-logistic	Generalised gamma	Good statistical and visual fit. All curves provide similar AUC. Consistent with PFS. Consistent with all risk population
TTP	Log-logistic	Weibull	Good statistical and visual fit. All feasible curves provide similar AUC. Consistency with PFS selection. Consistent with all risk population.
OS	Exponential	Weibull	Good statistical and visual fit. Midrange estimate. Consistent with all risk population.
Costs			
Number of administrations for fixed duration treatments based on	TTD	Mean number of administrations	Testing impact of using trial data on mean duration where available
RDI	Applied	Not applied	Data taken from numerous sources and uncertain whether or not IV therapies missed still incur a cost
Utilities			
Data source used for utilities	JAVELIN Renal 101 for 1L, AXIS trial for 2L and assumed same proportional decrease for 3L and 4L	CheckMate 9ER for 1L	CheckMate 9ER utilities higher than literature
		CheckMate 9ER for all lines	Fully aligned to trial utilities but these are higher than literature
		Same PFS and PD from 2L onwards	Data uncertain after 2L
		Higher proportional decrease for 3L and 4L	Decrease after 3L unclear
Utilities for BSC	Assumed same as progressed: current line	Assumed same as progression free: current line	Testing impact of alternative utilities for BSC
		Assume same as final health state	
Adverse events			

Parameter	Base case	Scenario	Justification
Data source used for AEs	NMA	Individual trials	Exploring impact of assuming same relative effectiveness for all G3+ AEs
AEs applied	One off	Per cycle	Exploring impact on how AE rates change over time
AE disutilities not considered	Yes	No	Potential for double counting in trial data sources
Scale of impact	Per analysis	Doubled	Clinical advice that AE impacts were underestimated in prior appraisals
Axitinib	Per NMA	Set the same as tivozanib	The NMA is based upon a small low-quality trial and clinical advice was that tivozanib and axitinib would be expected to have a similar safety profile given their similar MoA
Subsequent treatments			
Data source	RWE	Trial	Testing impact of subs therapy assumptions

Abbreviations: FP, fractional polynomial; MoA, mechanism of action; NMA, network meta-analysis; PARTSA, partitioned survival analysis; PH, proportional hazards; PFS, progression free survival; RDI, relative dosing intensity; RWE, real world evidence; TA, technology appraisal; TTD, time to discontinuation; UK, United Kingdom

5. COST-EFFECTIVENESS RESULTS

5.1. Model validation and face validity check

Within the model results and validation addendum which will follow this report, model outputs will be compared to the data used as model inputs (for example visual comparison to Kaplan Meier data) to ensure the appropriateness of model structure and data derivation. The model will then be compared to the projections from other models previously used for NICE STAs in the same decision point. Clinical expert input will be used to ensure that the model retains clinical face validity.

5.2. Benefits not captured in the QALY calculation

The only potential benefit identified that could not be included within the QALY calculation, is the potential benefit of cabozantinib within the combination for patients with bone metastases which was raised by one of the experts consulted. Literature, however, is conflicting as to whether or not there may be additional benefit in this subgroup.[265-267]

6. DISCUSSION AND CONCLUSIONS

6.1. Discussion

The major considerations identified for this appraisal include:

- Modelling methods, and outcomes of the cost-effectiveness analyses of various combinations, vary across the available literature including within prior NICE TAs. This underlines the benefit of a common modelling framework as far as practicable to enable consistency of decision making using the best available data at the time.
- Comparators for cabozantinib + nivolumab differ by risk status (combination therapies are only available outside of the CDF for intermediate / poor risk), which necessitates comparison by risk status; data for favourable risk patients is less well reported but what is available demonstrates that risk group is a potential treatment-effect modifier for IO/TKI combinations.
- Earlier treatment options affect what is available at later lines and may also impact on outcomes at later lines; data to be able to model the latter impact appears to be limited and prior appraisals have failed to meet Committee preferences to use UK data for the type of subsequent therapy received and to match costs and effectiveness.
- The outcomes demonstrated with RCTs showed greater absolute benefit than those demonstrated in SACT in a previous appraisal, indicating that use of RCT data for baseline risk may lead to an overestimate of benefit for treatments. This was also the case when comparing the RWE identified by the EAG in this pilot to the trials.
- The assumption of proportional hazards may not hold within RCC, but fractional polynomial NMAs pose additional challenges relating to estimability.
- Relatedly, the duration of treatment effect for newer combination treatments is uncertain, and evidence from a range of trials suggests ‘slippage’ in OS and PFS estimates with longer follow-up, particularly for IO/TKI combinations.
- NMAs broadly suggest that cabozantinib and nivolumab is an effective treatment in first line, but for intermediate and poor risk patients specific, long-term benefits against other treatments (including cabozantinib monotherapy) are less clear.
- NMAs at second line are challenged by difficulties linking networks to include all treatments.
- Sparseness of networks precluded exploration of key effect modifiers, though the EAG regarded that NMAs were feasible.
- There remain outstanding questions about the relevance of evidence across histologies, and the role of adjuvant pembrolizumab in impacting first-line treatment effectiveness.
- Our general modelling approach represents a shift from partitioned survival models to state transition models, though we preserve functionality for partitioned survival models. This ‘return’ to state transition models is necessary in order to have the flexibility to meet NICE’s objective to create a model capable of looking at the entire treatment pathway, though it also adds additional challenges in obtaining appropriate data and ensuring the plausibility of predictions of OS.

6.2. Conclusions for the cabozantinib + nivolumab appraisal

- In relation to the decision problem, the EAG disagreed on the full range of appropriate comparators, the relevance of time to next treatment, and the importance of risk group-specific analyses.
- The EAG noted a number of potential issues with respect to generalisability of the trial, including high rates of treatment after progression, over-optimistic estimates of OS and PFS compared to real-world evidence, low numbers of UK patients and low use of nivolumab after sunitinib, but was satisfied that the trial presented evidence of effectiveness of cabozantinib plus nivolumab as compared to sunitinib across key outcomes.
- Within the trial, there is evidence of modification by risk group for key outcomes, with systematically lower benefits for OS and PFS seen with more favourable risk.

6.3. Planned further work

- Model results will be provided in a separate addendum following this report
- The following additional work is planned to occur during the technical engagement phase of the cabozantinib + nivolumab appraisal:
  - Internal and external QC of the economic model
- The following additional work is planned to occur during final phases of the pilot after the appraisal of cabozantinib + nivolumab:
  - Review of clinical effectiveness information focussing specifically on sequencing and the impact of previous treatment on effectiveness.
  - Tidy up and genericise the model code for public release.
  - Addition of a Shiny user interface phase prior to public release.
  - Programming and analysis of model outputs related specifically to sequencing.
  - Consideration of how the platform model could be used for alternative decision making frameworks.
  - Release of the open source version of the economic model.

References

1. Lee D, Muthukumar M, Robinson S, Lovell A, Coelho H, Matthews J, et al. Treatments for renal cell carcinoma [ID6186]. Final Analysis Plan. Exeter: Peninsula Technology Assessment Group (PenTAG), University of Exeter; 2023. Available from: https://www.nice.org.uk/guidance/gid-ta11186/documents/assessment-report-3.

2. Jansen JP, Incerti D, Linthicum MT. Developing open-source models for the US Health System: practical experiences and challenges to date with the Open-Source Value Project. Pharmacoeconomics. 2019;37(11):1313-20.

3. Innovation and Value Initiative (IVI). IVI-NSCLC Value Model. Part of the Open-Source Value Project. Alexandria, VA: IVI. Available from: https://thevalueinitiative.org/ivi-nsclc-value-model/ (accessed 9 Jan 2023).

4. Cancer Research UK. Kidney cancer. Types and grades. 2020 [updated 9 March 2020]. Available from: https://www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/types-grades (accessed 9 March 2023).

5. Cancer Research UK. Kidney cancer. Symptoms. 2020 [updated 21 February 2020]. Available from: https://www.cancerresearchuk.org/about-cancer/kidney-cancer/symptoms (accessed 17 March 2023).

6. Sachdeva K, Jana B, Curti B. Renal Cell Carcinoma. Medscape; 2021. (accessed 17 March 2023).

7. Cella D. Beyond traditional outcomes: improving quality of life in patients with renal cell carcinoma. Oncologist. 2011;16(Suppl 2):23-31.

8. National Cancer Institute. Metastatic Cancer: When Cancer Spreads. 2020 [updated 10 November 2020]. Available from: https://www.cancer.gov/types/metastatic-cancer (accessed 17 March 2023).

9. Bianchi M, Sun M, Jeldres C, Shariat SF, Trinh QD, Briganti A, et al. Distribution of metastatic sites in renal cell carcinoma: a population-based analysis. Ann Oncol. 2012;23(4):973-80.

10. Cancer Research UK. Number stages for kidney cancer. 2020 [updated 3 March 2020]. Available from: https://www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/number-stages (accessed 3 January 2023).

11. Cancer Research UK. Kidney cancer incidence statistics. 2021 [updated 28 September 2021]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer/incidence (accessed 22 March 2023).

12. World Cancer Research Fund. Kidney Cancer. 2022. Available from: https://www.wcrf-uk.org/cancer-types/kidney-cancer/.

13. Cancer Research UK. Kidney cancer risk. 2019 [updated 18 April 2019]. (accessed 14 April 2023).

14. Public Health England. Cancer registration statistics, England: final release, 2018. London: Public Health England; 2020 [updated 29 May 2020]. Available from: https://www.gov.uk/government/statistics/cancer-registration-statistics-england-2018-final-release.

15. Office for National Statistics (ONS). Cancer survival in England: adult, stage at diagnosis and childhood - patients followed up to 2018. 2019 [updated 12 August 2019]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsandd

iseases/bulletins/cancersurvivalinengland/stageatdiagnosisandchildhoodpatientsfollowedupto2018.

16. Cancer Research UK. Kidney Cancer. Stages, types and grades. 2020 [updated 9 March 2020]. Available from: https://www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/types-grades (accessed 17 March 2023).

17. Kidney Care UK. Kidney Cancer. 2023. Available from: https://www.kidneycareuk.org/about-kidney-health/conditions/kidney-cancer/ (accessed 17 March 2023).

18. Cancer.Net. Kidney Cancer: Stages. 2021 [updated October 2021]. Available from: https://www.cancer.net/cancer-types/kidney-cancer/stages (accessed 20 April 2023).

19. National Cancer Registration and Analysis Service (NCRAS) [Internet]. National Disease Registration Service. 2023. Available from: https://www.cancerdata.nhs.uk/.

20. Bukowski RM. Prognostic factors for survival in metastatic renal cell carcinoma. Cancer. 2009;115(S10):2273-81.

21. Bhat S. Role of surgery in advanced/metastatic renal cell carcinoma. Indian J Urol. 2010;26(2):167-76.

22. Wagstaff J, Jones R, Hawkins R, Porfiri E, Pickering L, Bahl A, et al. Treatment patterns and clinical outcomes in patients with renal cell carcinoma in the UK: insights from the RECCORD registry. Ann Oncol. 2016;27(1):159-65.

23. Mani S, Todd MB, Katz K, Poo WJ. Prognostic factors for survival in patients with metastatic renal cancer treated with biological response modifiers. J Urol. 1995;154(1):35-40.

24. Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. J Clin Oncol. 2002;20(1):289-96.

25. Palmer PA, Vinke J, Philip T, Negrier S, Atzpodien J, Kirchner H, et al. Prognostic factors for survival in patients with advanced renal cell carcinoma treated with recombinant interleukin-2. Ann Oncol. 1992;3(6):475-80.

26. Landonio G, Baiocchi C, Cattaneo D, Ferrari M, Gottardi O, Majno M, et al. Retrospective analysis of 156 cases of metastatic renal cell carcinoma: evaluation of prognostic factors and response to different treatments. Tumori. 1994;80(6):468-72.

27. Heng DYC, Xie W, Regan MM, 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. J Clin Oncol. 2009;27(34):5794-9.

28. Heng DYC, Xie W, Regan MM, et al. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. Lancet Oncol. 2013;14(2):141-8.

29. George D, Jonasch E. Systemic therapy of advanced clear cell renal carcinoma. Waltham, MA: UpToDate; 2023. Available from: https://www.uptodate.com/contents/systemic-therapy-of-advanced-clear-cell-renal-carcinoma (accessed 5 January 2023).

30. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. J Clin Oncol. 1999;17(8):2530-40.

31. Okita K, Tanaka T, Fujita N, Hatakeyama S. Effect of Risk Classification Discrepancy Between the MSKCC and IMDC Models on Prognosis in Patients with Metastatic Renal Cell Carcinoma - Beyond the Abstract. UroToday. 2019 June 20.

32. Hawkins R, Fife K, Hurst M, Wang M, Naicker N, Nolasco S, et al. Treatment patterns and health outcomes in metastatic renal cell carcinoma patients treated with targeted systemic therapies in the UK. BMC Cancer. 2020;20(1):670.

33. Gore ME, Szczylik C, Porta C, Bracarda S, Bjarnason GA, Oudard S, 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.

34. Schmidinger M, Porta C, Oudard S, Denechere G, Brault Y, Serfass L, et al. Real-world Experience With Sunitinib Treatment in Patients With Metastatic Renal Cell Carcinoma: Clinical Outcome According to Risk Score. Clin Genitourin Cancer. 2020;18(5):e588-e97.

35. Yang YH, Ansel S, Meerveld-Eggink A, Jackson-Spence F, Rallis K, Brian P, et al. Baseline and Dynamic Changes in Hemoglobin Levels Predict Treatment Response and Prognosis in Metastatic Renal Cell Carcinoma: A Multicenter Retrospective Study. Clin Genitourin Cancer. 2023.

36. Yip S, Wells C, Moreira RB, Wong A, Srinivas S, Beuselinck B, et al. Real world experience of immuno-oncology agents in metastatic renal cell carcinoma: Results from the IMDC. Clin Oncol. 2017;35(6_suppl):492-.

37. NHS England. National Cancer Drugs Fund list - version 1.256. London2023 [updated 29 March 2023]. Available from: https://www.england.nhs.uk/publication/national-cancer-drugs-fund-list/ (accessed 31 March 2023).

38. National Institute for Health and Care Excellence (NICE). Lenvatinib with pembrolizumab for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA858]. London: NICE; 2023. Available from: https://www.nice.org.uk/guidance/ta858.

39. National Institute for Health and Care Excellence (NICE). Pembrolizumab for adjuvant treatment of renal cell carcinoma. Technology appraisal guidance [TA830]. London: NICE; 2022. Available from: https://www.nice.org.uk/guidance/ta830.

40. Larroquette M, Peyraud F, Domblides C, Lefort F, Bernhard J-C, Ravaud A, et al. Adjuvant therapy in renal cell carcinoma: Current knowledges and future perspectives. Cancer Treatment Reviews. 2021;97:102207.

41. Speed JM, Trinh QD, Choueiri TK, Sun M. Recurrence in Localized Renal Cell Carcinoma: a Systematic Review of Contemporary Data. Curr Urol Rep. 2017;18(2):15.

42. Pessoa RR, Kim S. BMJ Best Practice. Renal cell carcinoma. 2022 [updated 26 July 2022. Last reviewed 3 December 2022]. Available from: https://bestpractice.bmj.com/topics/en-gb/261 (accessed 3 January 2023).

43. Powles T, Albiges L, Bex A, et al. ESMO Clinical Practice Guideline update on the use of immunotherapy inearly stage and advanced renal cell carcinoma. Ann Oncol. 2021;32(12):1511-9.

44. Rathmell WK, Rumble RB, Veldhuizen PJV, Al-Ahmadie H, Emamekhoo H, Hauke RJ, et al. Management of Metastatic Clear Cell Renal Cell Carcinoma: ASCO Guideline. J Clin Oncol. 2022;40(25):2957-95.

45. Motzer R, Alekseev B, Rha SY, Porta C, Eto M, Powles T, et al. Lenvatinib plus Pembrolizumab or Everolimus for Advanced Renal Cell Carcinoma. N Engl J Med. 2021;384(14):1289-300.

46. National Institute for Health and Care Excellence (NICE). Avelumab with axitinib for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA645]. London: NICE; 2020. Available from: https://www.nice.org.uk/guidance/ta645.

47. National Institute for Health and Care Excellence (NICE). Pembrolizumab with axitinib for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA650]. London: NICE; 2020. Available from: https://www.nice.org.uk/guidance/TA650.

48. National Institute for Health and Care Excellence (NICE). Nivolumab with ipilimumab for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA780]. London2022. Available from: https://www.nice.org.uk/guidance/ta780.

49. National Institute for Health and Care Excellence (NICE). Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma. Technology appraisal guidance [TA169]. London: NICE; 2009. Available from: https://www.nice.org.uk/Guidance/TA169.

50. National Institute for Health and Care Excellence (NICE). Pazopanib for the first-line treatment of advanced renal cell carcinoma. Technology appraisal guidance [TA215]. London: NICE; 2011 [updated 1 August 2013]. Available from: https://www.nice.org.uk/Guidance/TA215.

51. National Institute for Health and Care Excellence (NICE). Tivozanib for treating advanced renal cell carcinoma. Technology appraisal guidance [TA512]. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ta512.

52. National Institute for Health and Care Excellence (NICE). Cabozantinib for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA542]. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ta542.

53. Challapalli A, Ratnayake G, Baht A, McGrane J, Frazer R. UK RCC RWE data set. Unpublished.

54. National Institute for Health and Care Excellence (NICE). Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment. Technology appraisal guidance [TA333]. London: NICE; 2015. Available from: https://www.nice.org.uk/guidance/ta333.

55. National Institute for Health and Care Excellence (NICE). Cabozantinib for previously treated advanced renal cell carcinoma. Technology appraisal guidance [TA463]. London: NICE; 2017. Available from: https://www.nice.org.uk/guidance/ta463.

56. National Institute for Health and Care Excellence (NICE). Lenvatinib with everolimus for previously treated advanced renal cell carcinoma. Technology appraisal guidance [TA498]. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ta498.

57. National Institute for Health and Care Excellence (NICE). Everolimus for advanced renal cell carcinoma after previous treatment. Technology appraisal guidance [TA432]. London: NICE; 2017. Available from: https://www.nice.org.uk/guidance/ta463.

58. National Institute for Health and Care Excellence (NICE). Nivolumab for previously treated advanced renal cell carcinoma. Technology appraisal guidance [TA417]. London: NICE; 2016. Available from: https://www.nice.org.uk/guidance/ta417.

59. Choueiri T, Powles T, Burotto M, et al. Nivolumab plus Cabozantinib versus Sunitinib for Advanced Renal-Cell Carcinoma. N Engl J Med. 2021;384(9):829-41.

60. EMA Recommends Extension of Therapeutic Indications for Cabozantinib and Nivolumab [press release]. Lugano: European Society for Medical Oncology, 04 March 2021 2021. Available from: https://www.esmo.org/oncology-news/ema-recommends-extension-of-therapeutic-indications-for-cabozantinib-and-nivolumab.

61. Medicines & Healthcare Products Regulatory Agency (MHRA). Summary of product characteristics. Cabometyx (cabozantinib) London: MHRA; 2021. Available from: https://mhraproducts4853.blob.core.windows.net/docs/66e37aeadd4825de7e965715964f804e632c92d0.

62. European Medicines Agency (EMA). Cabometyx. Summary of product characteristics. Amsterdam: European Medicines Agency; 2022 [updated 5 December 2022]. Available from: https://www.ema.europa.eu/en/documents/product-information/cabometyx-epar-product-information_en.pdf.

63. Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. Mol Cancer. 2018;17(1):58.

64. Xie Z, Lee YH, Boeke M, Jilaveanu LB, Liu Z, Bottaro DP, et al. MET Inhibition in Clear Cell Renal Cell Carcinoma. J Cancer. 2016;7(10):1205-14.

65. Zhou L, Liu XD, Sun M, Zhang X, German P, Bai S, et al. Targeting MET and AXL overcomes resistance to sunitinib therapy in renal cell carcinoma. Oncogene. 2016;35(21):2687-97.

66. Harshman LC, Choueiri TK. Targeting the hepatocyte growth factor/c-Met signaling pathway in renal cell carcinoma. Cancer J. 2013;19(4):316-23.

67. Yakes FM, Chen J, Tan J, Yamaguchi K, Shi Y, Yu P, et al. Cabozantinib (XL184), a novel MET and VEGFR2 inhibitor, simultaneously suppresses metastasis, angiogenesis, and tumor growth. Mol Cancer Ther. 2011;10(12):2298-308.

68. Rassy EE, Khoury Abboud RM, Ibrahim N, Assi T, Aoun F, Kattan J. The current state of immune checkpoint inhibitors in the first-line treatment of renal cancer. Immunotherapy. 2018;10(12):1047-52.

69. MHRA. Summary of product characteristics. Cabometyx (cabozantinib). Available from: https://mhraproducts4853.blob.core.windows.net/docs/016387065c25321d52c66e6d2d63ea8816104941 (accessed 28 March).

70. MHRA. Summary of product characteristics. Opdivo (nivolumab). Available from: https://mhraproducts4853.blob.core.windows.net/docs/8355d7f28eb4601b4f0cfd41f6b423cc0ecbb7ac (accessed 28 March).

71. BNF NICE (Cabozantinib). (accessed 17 March).

72. BNF NICE (Nivolumab). Available from: https://bnf.nice.org.uk/drugs/nivolumab/medicinal-forms/ (accessed 30 March).

73. Centre for Reviews and Dissemination. Systematic Reviews. CRD's guidance for undertaking reviews in health care. York: University of York; 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.

74. National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual. Process and methods [PMG36]. London: NICE; 2022. Available from: https://www.nice.org.uk/process/pmg36.

75. CADTH Search Filters Database. Ottawa: CADTH; 2022. Available from: https://searchfilters.cadth.ca.

76. Heo JH, Park C, Ghosh S, Park SK, Zivkovic M, Rascati KL. A network meta-analysis of efficacy and safety of first-line and second-line therapies for the management of metastatic renal cell carcinoma. J Clin Pharm Ther. 2021;46(1):35-49.

77. Liao Y, Hou H, Han Z, Liu Y. Systemic therapies for metastatic renal cell carcinoma in the second-line setting: A systematic review and network meta-analysis. Medicine. 2022;101(37):e30333.

78. Riaz IB, He H, Ryu AJ, Siddiqi R, Naqvi SAA, Yao Y, et al. A Living, Interactive Systematic Review and Network Meta-analysis of First-line Treatment of Metastatic Renal Cell Carcinoma[Formula presented]. Eur Urol. 2021;80(6):712-23.

79. Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan-Meier curves. Stata J. 2017;17(4):786-802.

80. National Institute for Health and Care Excellence (NICE). Single technology appraisal and highly specialised technologies evaluation: User guide for company evidence submission template. Process and methods [PMG24]. London: NICE; 2015 [updated 10 Feb 2022]. Available from: https://www.nice.org.uk/process/pmg24.

81. National Institute for Health and Care Excellence (NICE). NICE real-world evidence framework. Corporate document [ECD9]. London: NICE; 2022. Available from: https://www.nice.org.uk/corporate/ecd9/chapter/overview.

82. Scottish Intercollegiate Guidelines Network (SIGN). Search filters. Available from: https://www.sign.ac.uk/what-we-do/methodology/search-filters/ (accessed 9 January 2023).

83. Ayiku L, Levay P, Hudson T, Craven J, Barrett E, Finnegan A, et al. The medline UK filter: development and validation of a geographic search filter to retrieve research about the UK from OVID medline. Health Info Libr J. 2017;34(3):200-16.

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

85. Rini BI, Escudier B, Tomczak P, Kaprin A, Szczylik C, Hutson TE, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. Lancet. 2011;378(9807):1931-9.

86. Grünwald V, Hilser T, Meiler J, Goebell PJ, Ivanyi P, Strauss A, et al. A Prospectivly Randomized Phase-II Trial of Axitinib versus Everolimus as Second-Line Therapy in Metastatic Renal Cell Carcinoma (BERAT Study). Oncol Res Treat. 2022;45(5):272-80.

87. Vano YA, Elaidi R, Bennamoun M, Chevreau C, Borchiellini D, Pannier D, et al. Nivolumab, nivolumab-ipilimumab, and VEGFR-tyrosine kinase inhibitors as first-line treatment for metastatic clear-cell renal cell carcinoma (BIONIKK): a biomarker-driven, open-label, non-comparative, randomised, phase 2 trial. Lancet Oncol. 2022;23(5):612-24.

88. Choueiri TK, Hessel C, Halabi S, Sanford B, Michaelson MD, Hahn O, 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. Eur J Cancer. 2018;94:115-25.

89. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. N Engl J Med. 2015;373(19):1803-13.

90. Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. N Engl J Med. 2018;378(14):1277-90.

91. Motzer RJ, Hutson TE, Cella D, Reeves J, Hawkins R, Guo J, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. N Engl J Med. 2013;369(8):722-31.

92. Tomita Y, Naito S, Sassa N, Takahashi A, Kondo T, Koie T, et al. Sunitinib Versus Sorafenib as Initial Targeted Therapy for mCC-RCC With Favorable/Intermediate Risk: Multicenter Randomized Trial CROSS-J-RCC. Clin Genitourin Cancer. 2020;18(4):e374-e85.

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

94. Hutson TE, Lesovoy V, Al-Shukri S, Stus VP, Lipatov ON, Bair AH, et al. Axitinib versus sorafenib as first-line therapy in patients with metastatic renal-cell carcinoma: a randomised open-label phase 3 trial. Lancet Oncol. 2013;14(13):1287-94.

95. Motzer RJ, Penkov K, Haanen J, Rini B, Albiges L, Campbell MT, et al. Avelumab plus Axitinib versus Sunitinib for Advanced Renal-Cell Carcinoma. N Engl J Med. 2019;380(12):1103-15.

96. Choueiri TK, Escudier B, Powles T, Mainwaring PN, Rini BI, Donskov F, et al. Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. N Engl J Med. 2015;373(19):1814-23.

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

98. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. Lancet. 2008;372(9637):449-56.

99. Motzer RJ, Barrios CH, Kim TM, Falcon S, Cosgriff T, Harker WG, et al. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. J Clin Oncol. 2014;32(25):2765-72.

100. Eichelberg C, Vervenne WL, De Santis M, Fischer von Weikersthal L, Goebell PJ, Lerchenmüller C, et al. SWITCH: A Randomised, Sequential, Open-label Study to Evaluate the Efficacy and Safety of Sorafenib-sunitinib Versus Sunitinib-sorafenib in the Treatment of Metastatic Renal Cell Cancer. Eur Urol. 2015;68(5):837-47.

101. Retz M, Bedke J, Bögemann M, Grimm MO, Zimmermann U, Müller L, et al. SWITCH II: Phase III randomized, sequential, open-label study to evaluate the efficacy and safety of sorafenib-pazopanib versus pazopanib-sorafenib in the treatment of advanced or metastatic renal cell carcinoma (AUO AN 33/11). Eur J Cancer. 2019;107:37-45.

102. Pal SK, Tangen C, Thompson IM, Jr., Balzer-Haas N, George DJ, Heng DYC, et al. A comparison of sunitinib with cabozantinib, crizotinib, and savolitinib for treatment of advanced papillary renal cell carcinoma: a randomised, open-label, phase 2 trial. Lancet. 2021;397(10275):695-703.

103. Motzer RJ, Nosov D, Eisen T, Bondarenko I, Lesovoy V, Lipatov O, et al. Tivozanib versus sorafenib as initial targeted therapy for patients with metastatic renal cell carcinoma: results from a phase III trial. J Clin Oncol. 2013;31(30):3791-9.

104. Rini BI, Pal SK, Escudier BJ, Atkins MB, Hutson TE, Porta C, et al. Tivozanib versus sorafenib in patients with advanced renal cell carcinoma (TIVO-3): a phase 3, multicentre, randomised, controlled, open-label study. Lancet Oncol. 2020;21(1):95-104.

105. Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. J Clin Oncol. 2010;28(6):1061-8.

106. Motzer RJ, McDermott DF, Escudier B, Burotto M, Choueiri TK, Hammers HJ, et al. Conditional survival and long-term efficacy with nivolumab plus ipilimumab versus sunitinib in patients with advanced renal cell carcinoma. Cancer. 2022;128(11):2085-97.

107. Motzer RJ, Escudier B, George S, Hammers HJ, Srinivas S, Tykodi SS, et al. Nivolumab versus everolimus in patients with advanced renal cell carcinoma: Updated results with long-term follow-up of the randomized, open-label, phase 3 CheckMate 025 trial. Cancer. 2020;126(18):4156-67.

108. Ahrens M, Escudier B, Haanen J, Boleti E, Gross-Goupil M, Grimm MO, et al. An ongoing, randomized phase II study of Nivolumab plus Ipilimumab versus standard of care in previously untreated and advanced non-clear cell renal cell carcinoma (SUNIFORECAST). Oncol Res Treat. 2022;45(Supplement 3):100-1.

109. Motzer RJ, Porta C, Eto M, Powles T, Grünwald V, Alekseev B, et al. Final prespecified overall survival (OS) analysis of CLEAR: 4-year follow-up of lenvatinib plus pembrolizumab (L+P) vs sunitinib (S) in patients (pts) with advanced renal cell carcinoma (aRCC). J Clin Oncol. 2023;41(16_suppl):4502-.

110. Ipsen. Company submission for "Treatments for renal cell carcinoma [ID6186]: pathways pilot appraisal". Unpublished.

111. Pond GR. Statistical issues in the use of dynamic allocation methods for balancing baseline covariates. Br J Cancer. 2011;104(11):1711-5.

112. Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). Cochrane Methods; 2019 22 August 2019. Available from: https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2.

113. Haanen JBAG, Larkin J, Choueiri TK, Albiges L, Rini BI, Atkins MB, et al. Extended follow-up from JAVELIN Renal 101: subgroup analysis of avelumab plus axitinib versus sunitinib by the International Metastatic Renal Cell Carcinoma Database Consortium risk group in patients with advanced renal cell carcinoma. ESMO Open. 2023;8(3):101210.

114. Verzoni E, Escudier B, Hutson TE, McDermott DF, Pal SK, Porta C, et al. TIVO-3: Durability of response and updated overall survival of tivozanib versus sorafenib in metastatic renal cell carcinoma (mRCC). J Clin Oncol. 2021;39(15_suppl):4546-.

115. Cella D, Yount S, Brucker PS, Du H, Bukowski R, Vogelzang N, et al. Development and validation of a scale to measure disease-related symptoms of kidney cancer. Value Health. 2007;10(4):285-93.

116. Cella D, Motzer RJ, Rini BI, Cappelleri JC, Ramaswamy K, Hariharan S, et al. Important Group Differences on the Functional Assessment of Cancer Therapy-Kidney Symptom Index Disease-Related Symptoms in Patients with Metastatic Renal Cell Carcinoma. Value Health. 2018;21(12):1413-8.

117. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. Journal of Clinical Oncology. 1998;16(1):139-44.

118. 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(1):70.

119. Motzer RJ, Hutson TE, McCann L, Deen K, Choueiri TK. Overall survival in renal-cell carcinoma with pazopanib versus sunitinib. N Engl J Med. 2014;370(18):1769-70.

120. Burotto M, Powles T, Escudier B, Apolo AB, Bourlon MT, Shah AY, et al. Nivolumab plus cabozantinib vs sunitinib for first-line treatment of advanced renal cell carcinoma (aRCC): 3-year follow-up from the phase 3 CheckMate 9ER trial. J Clin Oncol. 2023;41(6_suppl):603-.

121. Cella D, Choueiri TK, Blum SI, Ejzykowicz F, Wallace J, Zhang J, et al. Health-related quality of life (HRQoL) of risk-based patient subgroups with advanced renal cell cancer (aRCC) treated with nivolumab plus cabozantinib (NIVO+CABO) vs sunitinib (SUN) in the CheckMate 9ER trial. J Clin Oncol. 2023;41(16_suppl):4527-.

122. Cella D, Motzer RJ, Suarez C, Blum SI, Ejzykowicz F, Hamilton M, et al. Patient-reported outcomes with first-line nivolumab plus cabozantinib versus sunitinib in patients with advanced renal cell carcinoma treated in CheckMate 9ER: an open-label, randomised, phase 3 trial. Lancet Oncol. 2022;23(2):292-303.

123. Ljungberg B, Albiges L, Abu-Ghanem Y, Bedke J, Capitanio U, Dabestani S, et al. European Association of Urology Guidelines on Renal Cell Carcinoma: The 2022 Update. Eur Urol. 2022;82(4):399-410.

124. National Institute for Health and Care Excellence (NICE). Kidney Cancer In Development (GID-NG10398). Available at: https://www.nice.org.uk/guidance/indevelopment/gid-ng10398 (last accessed June 2023).

125. SACT Systemic Anti-Cancer Therapy. Chemotherapy Dataset [Internet]. Public Health England. 2022. Available from: http://www.chemodataset.nhs.uk/home.

126. Clinical Practice Research Datalink (CPRD) [Internet]. The Medicines and Healthcare products Regulatory Agency. 2023. Available from: https://cprd.com/.

127. Hospital Episode Statistics (HES) [Internet]. NHS Digital. 2023. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.

128. Marchioni M, Kriegmair M, Heck M, Amiel T, Porpiglia F, Ceccucci E, et al. Development of a Novel Risk Score to Select the Optimal Candidate for Cytoreductive Nephrectomy Among Patients with Metastatic Renal Cell Carcinoma. Results from a Multi-institutional Registry (REMARCC). Eur Urol Oncol. 2021;4(2):256-63.

129. International mRCC Database Consortium [Internet]. 2021. Available from: https://www.imdconline.com/.

130. Maroun R, Mitrofan L, Benjamin L, Nachbaur G, Maunoury F, Le Jeunne P, et al. Real life patterns of care and progression free survival in metastatic renal cell carcinoma patients: retrospective analysis of cross-sectional data. BMC Cancer. 2018;18(1):214.

131. Olsson-Brown AC, Baxter M, Dobeson C, Feeney L, Lee R, Maynard A, et al. Real-world outcomes of immune-related adverse events in 2,125 patients managed with immunotherapy: A United Kingdom multicenter series. J Clin Oncol. 2020;38(15_suppl):7065-.

132. Nathan PD, Muazzam IA, Frazer R, Sharma A, Hickey JD, Ritchie AR. Avelumab plus axitinib in advanced renal cell carcinoma (aRCC): 12-month interim results from a real-world observational study in the United Kingdom. J Clin Oncol. 2022;40(6_suppl):301-.

133. Brown JE, Harrow B, Marciniak A, McCarthy C, Cirneanu L, Protheroe A. Cabozantinib and axitinib after VEGF therapy in patients with aRCC: A retrospective cohort study. Ann Oncol. 2021;32(Supplement 5):S701.

134. Hack J, Hill SL, Broadfoot J, Chau C, Geldart T, Gale J, et al. Real world experience of nivolumab therapy in metastatic renal cancer patients: A 3 year multi-centre review. Ann Oncol. 2019;30(Supplement 5):v389.

135. Kidney Cancer UK: Quality Performance Audit of kidney cancer services in England Cambridge: Kidney Cancer UK; 2022. Available from: https://www.kcuk.org.uk/wp-content/uploads/2022/08/Kidney-Cancer-UK-Accord-AUDIT-REPORT-VFINAL.pdf.

136. Nathan PD, Allison J, Charnley N, Michael A, Moore K, Sharma A, et al. CARINA interim analysis: A non-interventional study of real-world treatment sequencing and outcomes in patients with advanced renal cell carcinoma initiated on first-line checkpoint inhibitor-based combination therapy. J Clin Oncol. 2023;41(6_suppl):626-.

137. IQVIA. Hospital Audit Data 2022. Unpublished.

138. Hilser T, Darr C, Niegisch G, Schnabel MJ, Foller S, Haeuser L, et al. Cabozantinib + nivolumab in adult patients with advanced or metastatic renal cell carcinoma: A retrospective, non-interventional study in a real-world cohort. J Clin Oncol. 2023;41(6_suppl):628-.

139. Challapalli A, Ratnayake G, McGrane J, Frazer R, Gupta S, Parslow DS, et al. 1463P Patterns of care and outcomes of metastatic renal cell carcinoma (mRCC) patients (pts) with bone metastases (BM): A UK multicenter review. Ann Oncol. 2022;33(Supplement 7):S1215.

140. Santoni M, Heng DY, Bracarda S, Procopio G, Milella M, Porta C, et al. Real-World Data on Cabozantinib in Previously Treated Patients with Metastatic Renal Cell Carcinoma: Focus on Sequences and Prognostic Factors. Cancers (Basel). 2019;12(1).

141. Dias S, Ades A, Welton N, Jansen J, Sutton A. Network Meta-Analysis for Decision Making. Hoboken, NJ: John Wiley & Sons Ltd; 2018.

142. Freeman SC, Cooper NJ, Sutton AJ, Crowther MJ, Carpenter JR, Hawkins N. Challenges of modelling approaches for network meta-analysis of time-to-event outcomes in the presence of non-proportional hazards to aid decision making: Application to a melanoma network. Stat Methods Med Res. 2022;31(5):839-61.

143. Wiksten A, Hawkins N, Piepho H-P, Gsteiger S. Nonproportional Hazards in Network Meta-Analysis: Efficient Strategies for Model Building and Analysis. Value Health. 2020;23(7):918-27.

144. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med. 2015;34(6):984-98.

145. Gsteiger S, Howlett N, Ashlee B. gemtcPlus: Provides a suite of extension functions for NMA using the 'gemtc' package. R package version 1.0.0. . F. Hoffmann La Roche Ltd; 2022. Available from: https://roche.github.io/gemtcPlus/.

146. Collett D. Modelling Survival Data in Medical Research. 3rd ed. New York (NY): Chapman and Hall/CRC; 2014.

147. EU Clinical Trials Register. A prospective, open-label, multicenter, randomized phase II trial: Sequential therapy with BEvacizumab, RAd001 (everolimus) and AxiTinib in metastatic renal cell carcinoma (mRCC) (BERAT study). EudraCT Number: 2011-005939-78. Amsterdam: European Medicines Agency; 2012. Available from: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2011-005939-78 (accessed 28 June 2023).

148. Cope S, Zhang J, Saletan S, Smiechowski B, Jansen JP, Schmid P. A process for assessing the feasibility of a network meta-analysis: a case study of everolimus in combination with hormonal therapy versus chemotherapy for advanced breast cancer. BMC Med. 2014;12:93.

149. Klaassen Z. ASCO 2023: Final Prespecified Overall Survival Analysis of CLEAR: 4-Year Follow-up of Lenvatinib + Pembrolizumab vs Sunitinib in Patients with Advanced RCC. UroToday; 2023. Available from: https://www.urotoday.com/conference-highlights/asco-2023/asco-2023-kidney-cancer/144971-asco-2023-final-prespecified-overall-survival-analysis-of-clear-4-year-follow-up-of-lenvatinib-pembrolizumab-vs-sunitinib-in-patients-with-advanced-rcc.html.

150. Aldin A, Besiroglu B, Adams A, Monsef I, Piechotta V, Tomlinson E, et al. First-line therapy for adults with advanced renal cell carcinoma: a systematic review and network meta-analysis. Cochrane Database Syst Rev. 2023;5(5):Cd013798.

151. ClinicalTrials.gov. Lenvatinib With Everolimus Versus Cabozantinib for Second-Line or Third-Line Treatment of Metastatic Renal Cell Cancer. ClinicalTrials.gov Identifier: NCT05012371. 2021 [updated 26 January 2023]. Available from: https://clinicaltrials.gov/ct2/show/NCT05012371 (accessed 27 March 2023).

152. ClinicalTrials.gov. Randomized Phase-II Study of Nivolumab Plus Ipilimumab vs. Standard of Care in Untreated and Advanced Non-clear Cell RCC. ClinicalTrials.gov

Identifier: NCT03075423. 2017 [updated 3 November 2022]. Available from: https://clinicaltrials.gov/ct2/show/NCT03075423 (accessed 27 March 2023).

153. ClinicalTrials.gov. A Study to Compare Treatments for a Type of Kidney Cancer Called TFE/Translocation Renal Cell Carcinoma (tRCC). ClinicalTrials.gov Identifier: NCT03595124. 2018 [updated 23 March 2023]. Available from: https://clinicaltrials.gov/ct2/show/NCT03595124 (accessed 27 March 2023).

154. ClinicalTrials.gov. Cabozantinib or Sunitinib Malate in Treating Participants With Metastatic Variant Histology Renal Cell Carcinoma. ClinicalTrials.gov Identifier: NCT03541902. 2018 [updated 14 December 2022]. Available from: https://clinicaltrials.gov/ct2/show/NCT03541902 (accessed 27 March 2023).

155. ClinicalTrials.gov. REduced Frequency ImmuNE Checkpoint Inhibition in Cancers (REFINE). ClinicalTrials.gov Identifier: NCT04913025. 2021 [updated 21 September 2022]. Available from: https://clinicaltrials.gov/ct2/show/NCT04913025.

156. ClinicalTrials.gov. A Study of Subcutaneous Nivolumab Monotherapy With or Without Recombinant Human Hyaluronidase PH20 (rHuPH20). ClinicalTrials.gov Identifier: NCT03656718. 2018 [updated 20 March 2023]. Available from: https://clinicaltrials.gov/ct2/show/NCT03656718.

157. ClinicalTrials.gov. Renal Adjuvant MultiPle Arm Randomised Trial (RAMPART). ClinicalTrials.gov Identifier: NCT03288532. 2017 [updated 7 September 2020]. Available from: https://clinicaltrials.gov/ct2/show/NCT03288532 (accessed 1 June 2023).

158. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess. 2004;8(36):1-158.

159. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Marshall C, et al. Chapter 4: Searching for and selecting studies. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022): Cochrane; 2022.

160. Woods B, Sideris E, Palmer S, Latimer N, Soares M. TSD 19: Partitioned survival analysis as a decision modelling tool. Sheffield: School of Health and Related Research, University of Sheffield; 2017. Available from: https://www.sheffield.ac.uk/media/34205/download?attachment.

161. Verma V, Sprave T, Haque W, Simone CB, 2nd, Chang JY, Welsh JW, et al. A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors. J Immunother Cancer. 2018;6(1):128.

162. Philip EJ, Zhang S, Tahir P, Kim D, Wright F, Bell A, et al. Cost-Effectiveness of Immunotherapy Treatments for Renal Cell Carcinoma: A Systematic Review. Kidney Cancer. 2021;5(1):47-62.

163. Sarfaty M, Leshno M, Gordon N, Moore A, Neiman V, Rosenbaum E, et al. Cost Effectiveness of Nivolumab in Advanced Renal Cell Carcinoma. Eur Urol. 2018;73(4):628-34.

164. McCrea C, Johal S, Yang S, Doan J. Cost-effectiveness of nivolumab in patients with advanced renal cell carcinoma treated in the United States. Exp Hematol Oncol. 2018;7:4.

165. Wan XM, Peng LB, Ma JA, Li YJ. Economic evaluation of nivolumab as a second-line treatment for advanced renal cell carcinoma from US and Chinese perspectives. Cancer. 2017;123(14):2634-41.

166. Li S, Li J, Peng L, Li Y, Wan X. Cost-Effectiveness of Nivolumab Plus Cabozantinib Versus Sunitinib as a First-Line Treatment for Advanced Renal Cell Carcinoma in the United States. Front Pharmacol. 2021;12:736860.

167. Liao W, Lei W, Feng M, Yang Y, Wu Q, Zhou K, et al. Cost-Effectiveness Analysis of First-Line Nivolumab Plus Cabozantinib for Advanced Renal Cell Carcinoma in the United States. Adv Ther. 2021;38(12):5662-70.

168. Liu R, Qiu K, Wu J, Jiang Y, Wu P, Pang J. Cost-effectiveness analysis of nivolumab plus cabozantinib versus sunitinib as first-line therapy in advanced renal cell carcinoma. Immunotherapy. 2022;14(11):859-69.

169. Wang H, Wang Y, Li L, Zhou H, Lili S, Li L, et al. Economic evaluation of first-line nivolumab plus cabozantinib for advanced renal cell carcinoma in China. Front Public Health. 2022;10:954264.

170. Marciniak A, Gultyaev D, Orbzut G, Mollon P, Wallace JF. Cost Effectiveness of Cabozantinib Plus Nivolumab As First-Line Treatment for Renal Cell Carcinoma. Value Health. 2022;25(7 Supplement):S300-S1.

171. Tempelaar S, Kroep S, Marie L, Juban L, Kurt M, Ejzykowicz F, et al. POSA150 Cost-Effectiveness Analysis of First-Line (1L) Systemic Treatments in Advanced Renal Cell Carcinoma (aRCC) in France. Value Health. 2022;25(1 Supplement):S63.

172. Yoshida L, Serafini P, Picoli R, Castelano N, Senra F. EE130 Cost-Utility of Cabonivo vs Sunitinib, Pazopanib and Ipinivo for First Line Advanced Renal-Cell Carcinoma in the Private Healthcare System in Brazil. Value Health. 2022;25(12 Supplement):S78.

173. Marciniak A, Gultyaev D, Obrzut G, Molton P, Wallace JF, editors. Cost effectiveness of cabozantinib plus nivolumab as firstline treatment for renal cell carcinoma [presentation]. International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 2022 annual meeting May 1518, 2022 2022 15-18 May 2022; Washington, DC, USA.

174. Cella D, Escudier B, Tannir NM, Powles T, Donskov F, Peltola K, et al. Quality of Life Outcomes for Cabozantinib Versus Everolimus in Patients With Metastatic Renal Cell Carcinoma: METEOR Phase III Randomized Trial. J Clin Oncol. 2018;36(8):757-64.

175. de Groot S, Redekop WK, Versteegh MM, Sleijfer S, Oosterwijk E, Kiemeney LALM, et al. Health-related quality of life and its determinants in patients with metastatic renal cell carcinoma. Qual Life Res. 2018;27(1):115-24.

176. Wan X, Zhang Y, Tan C, Zeng X, Peng L. First-line Nivolumab Plus Ipilimumab vs Sunitinib for Metastatic Renal Cell Carcinoma: A Cost-effectiveness Analysis. JAMA Oncol. 2019;5(4):491-6.

177. Patel KK, Giri S, Parker TL, Bar N, Neparidze N, Huntington SF. Cost-Effectiveness of First-Line Versus Second-Line Use of Daratumumab in Older, Transplant-Ineligible Patients With Multiple Myeloma. J Clin Oncol. 2021;39(10):1119-28.

178. Wu B, Zhang Q, Sun J. Cost-effectiveness of nivolumab plus ipilimumab as first-line therapy in advanced renal-cell carcinoma. J Immunother Cancer. 2018;6(1):124.

179. Deniz B, Ambavane A, Yang S, Altincatal A, Doan J, Rao S, et al. Treatment sequences for advanced renal cell carcinoma: A health economic assessment. PloS One. 2019;14(8):e0215761.

180. Doleh Y, Deniz B, Ambavane A, Rao S, Page V, Michaelson MD. Health economic assessment of treatment sequences for the management of patients with metastatic renal cell carcinoma in the United States. Value Health. 2017;20(5):A309.

181. Mason NT, Joshi VB, Adashek JJ, Kim Y, Shah SS, Schneider AM, et al. Cost Effectiveness of Treatment Sequences in Advanced Renal Cell Carcinoma. Eur Urol Oncol. 2023.

182. De Groot S, Blommestein HM, Redekop WK, Sleijfer S, Kiemeney LALM, Oosterwijk E, et al. Potential health gains for patients with metastatic renal cell carcinoma in daily

clinical practice: A real-world cost-effectiveness analysis of sequential first- and second-line treatments. PloS One. 2017;12(5):e0177364.

183. Raphael J, Sun Z, Bjarnason GA, Sander B, Naimark DM. Nivolumab in the treatment of metastatic renal cell carcinoma: A cost-utility analysis. J Clin Oncol. 2017;35(15_suppl):e18331-e.

184. Raphael J, Sun Z, Bjarnason GA, Helou J, Sander B, Naimark DM. Nivolumab in the Treatment of Metastatic Renal Cell Carcinoma: A Cost-Utility Analysis. Am J Clin Oncol. 2018;41(12):1235-42.

185. Zhu Y, Liu K, Ding D, Peng L. First-Line Lenvatinib Plus Pembrolizumab or Everolimus versus Sunitinib for Advanced Renal Cell Carcinoma: A United States-Based Cost-Effectiveness Analysis. Clin Genitourin Cancer. 2023;21(3):417.e1-.e10.

186. Yfantopoulos N, Bafaloukos I, Dimitriadis I, Skroumpelos A, Karokis A. EE652 Cost Effectiveness Analysis of Pembrolizumab for the Treatment of Advanced or Metastatic Renal Cell Carcinoma. Value Health. 2022;25(12):S184-S5.

187. Nazha S, Tanguay S, Kapoor A, Jewett M, Kollmannsberger C, Wood L, et al. Cost-utility of Sunitinib Versus Pazopanib in Metastatic Renal Cell Carcinoma in Canada using Real-world Evidence. Clin Drug Investig. 2018;38(12):1155-65.

188. Escudier B, Motzer RJ, Dyer M, May JR, Ejzykowicz F, Kurt M, et al. 1459P Analysis of long-term efficacy outcomes from the CheckMate 025 (CM 025) trial comparing nivolumab (NIVO) vs everolimus (EVE) based on ≥ 7 years (yrs) of follow-up in pre-treated patients (pts) with advanced renal cell carcinoma (aRCC). Ann Oncol. 2022;33:S1212.

189. Bojke L, Soares M, Claxton K, Colson A, Fox A, Jackson C, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technol Assess. 2021;25(37):1-124.

190. University of York. Centre for Health Economics. Structured expert elicitation resources (STEER). Available from: https://www.york.ac.uk/che/research/teehta/elicitation/steer/ (accessed 9 January 2023).

191. Harding G, Cella D, Robinson D, Jr., Mahadevia PJ, Clark J, Revicki DA. Symptom burden among patients with renal cell carcinoma (RCC): content for a symptom index. Health Qual Life Outcomes. 2007;5:34.

192. Tenold M, Ravi P, Kumar M, Bowman A, Hammers H, Choueiri TK, et al. Current Approaches to the Treatment of Advanced or Metastatic Renal Cell Carcinoma. ASCO Educational Book. 2020(40):187-96.

193. McGrane J, Frazer R, Challapalli A, Ratnayake G, Lydon A, Parslow DS, et al. A multicenter real-world study reviewing systemic anticancer treatment choices and drop off rates between treatment lines for metastatic renal cell carcinoma in the United Kingdom: In the immunotherapy era. J Clin Oncol. 2022;40(6_suppl):358-.

194. Alam R, Patel HD, Su ZT, Cheaib JG, Ged Y, Singla N, et al. Self-reported quality of life as a predictor of mortality in renal cell carcinoma. Cancer. 2022;128(3):479-86.

195. Zhang J, Liang W, Liang H, Wang X, He J. Endpoint surrogacy in oncological randomized controlled trials with immunotherapies: a systematic review of trial-level and arm-level meta-analyses. Ann Transl Med. 2019;7(11):244.

196. Abdel-Rahman O. Surrogate end points for overall survival in trials of PD-(L)1 inhibitors for urinary cancers: a systematic review. Immunotherapy. 2018;10(2):139-48.

197. Johal S, Santi I, Doan J, George S. Is RECIST-defined progression free-survival a meaningful endpoint in the era of immunotherapy? J Clin Oncol. 2017;35(6_suppl):488-.

198. Johnson KR, Liauw W, Lassere MN. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: a systematic review. Ann Oncol. 2015;26(3):485-96.

199. Nie RC, Chen FP, Yuan SQ, Luo YS, Chen S, Chen YM, et al. Evaluation of objective response, disease control and progression-free survival as surrogate end-points for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials. Eur J Cancer. 2019;106:1-11.

200. Bria E, Massari F, Maines F, Pilotto S, Bonomi M, Porta C, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma. Correlation with overall survival, benchmarking and power analysis. Crit Rev Oncol Hematol. 2015;93(1):50-9.

201. Delea TE, Khuu A, Kay A, Zheng J, Baladi JF. Association between treatment effects on disease progression (DP) endpoints and overall survival (OS) in patients with metastatic renal cell carcinoma (mRCC). J Clin Oncol. 2009;27(15_suppl):5105-.

202. Neqrier S, Bushmakin A, Cappelleri J, Charbonneau C, Sandin R, Michaelson M, et al. 1301P Assessment of Progression-free Survival as a Surrogate Endpoint for Overall Survival in Patients With Metastatic Renal Cell Carcinoma. Eur J Cancer. 2011;47:S163.

203. Petrelli F, Barni S. Surrogate End Points and Postprogression Survival in Renal Cell Carcinoma: An Analysis of First-Line Trials With Targeted Therapies. Clin Genitourin Cancer. 2013;11(4):385-9.

204. Branchoux S, Sofeu CL, Kurt M, Gaudin AF, Italiano A, Rondeau V, et al. EX3 Investigating Time to Next Treatment As a Surrogate Endpoint for Overall Survival in Previously Untreated Intermediate- to Poor-Risk Advanced Renal Cell Carcinoma Patients: An Insight from the Phase III CheckMate-214. Value Health. 2020;23:S402.

205. Torres AZ, Nussbaum NC, Parrinello CM, Bourla AB, Bowser BE, Wagner S, et al. Analysis of a Real-World Progression Variable and Related Endpoints for Patients with Five Different Cancer Types. Adv Ther. 2022;39(6):2831-49.

206. Ismail RK, Schramel FMNH, van Dartel M, Pasmooij AMG, Cramer-van der Welle CM, Hilarius DL, et al. Individual patient data to allow a more elaborated comparison of trial results with real-world outcomes from second-line immunotherapy in NSCLC. BMC Med Res Methodol. 2023;23(1):1.

207. Zhu J, Tang RS. A proper statistical inference framework to compare clinical trial and real-world progression-free survival data. Stat Med. 2022;41(29):5738-52.

208. National Institute for Health and Care Excellence (NICE). Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated. Technology appraisal guidance [TA387]. London: NICE; 2016 [updated 27 July 2016]. Available from: https://www.nice.org.uk/guidance/ta387.

209. National Institute for Health and Care Excellence (NICE). Fedratinib for treating disease-related splenomegaly or symptoms in myelofibrosis. Technology appraisal guidance [TA756]. London: NICE; 2021. Available from: https://www.nice.org.uk/guidance/ta756.

210. National Institute for Health and Care Excellence (NICE). Ruxolitinib for treating disease-related splenomegaly or symptoms in adults with myelofibrosis. Technology appraisal guidance [TA386]. London: NICE; 2016. Available from: https://www.nice.org.uk/guidance/ta386.

211. National Institute for Health and Care Excellence (NICE). Ruxolitinib for disease-related splenomegaly or symptoms in adults with myelofibrosis (review of TA289) [ID831]. Committee papers. London: NICE; 2016. Available from: https://www.nice.org.uk/guidance/ta386/documents/committee-papers.

212. National Institute for Health and Care Excellence (NICE). Abiraterone acetate for the treatment of metastatic hormone relapsed prostate cancer not previously treated with chemotherapy. Premeeting briefing. London: NICE; 2014. Available from: https://www.nice.org.uk/guidance/ta387/documents/prostate-cancer-metastatic-hormone-relapsed-not-treated-with-chemotherapy-abiraterone-acetate-with-prednisolone-id503-evaluation-report2.

213. Kaltenthaler E, Tappenden P, Paisley S, Squires H. TSD 13: Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. Sheffield: School of Health and Related Research, University of Sheffield 2011. Available from: https://www.sheffield.ac.uk/media/34222/download?attachment.

214. Davis S, Stevenson M, Tappenden P, Wailoo A. TSD 15: Cost-effectiveness modelling using patient-level simulation. Sheffield: School of Health and Related Research, University of Sheffield 2014. Available from: https://www.sheffield.ac.uk/media/34206/download?attachment.

215. Motzer RJ, Powles T, Burotto M, et al. Nivolumab plus cabozantinib versus sunitinib in first-line treatment for advanced renal cell carcinoma (CheckMate 9ER): long-term follow-up results from an open-label, randomised, phase 3 trial. Lancet Oncol. 2022;23:888-98.

216. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. Med Decis Making. 2017;37(4):340-52.

217. National Institute for Health and Care Excellence (NICE). Durvalumab for maintenance treatment of unresectable non-small-cell lung cancer after platinum-based chemoradiation. Technology appraisal guidance [TA798]. London: NICE; 2022. Available from: https://www.nice.org.uk/guidance/ta798.

218. Incerti D, Thom H, Baio G, Jansen JP. R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment. Value Health. 2019;22(5):575-9.

219. Hart R, Burns D, Ramaekers B, Ren S, Gladwell D, Sullivan W, et al. R and Shiny for Cost-Effectiveness Analyses: Why and When? A Hypothetical Case Study. Pharmacoeconomics. 2020;38(7):765-76.

220. Zorginstituut Nederland. Richtlijn kosteneffectiviteitsmodellen in R [Cost-effectiveness models directive in R]. Diemen: Zorginstituut Nederland [Dutch Healthcare Institute]; 2022. Available from: https://www.zorginstituutnederland.nl/publicaties/publicatie/2022/12/15/richtlijn-kosteneffectiviteitsmodellen-in-r

221. Shiny. Easy web applications in R. Boston (MA): Posit Software; 2023. Available from: https://www.rstudio.com/products/shiny/.

222. European Medicines Agency (EMA). Votrient. Summary of product characteristics. Amsterdam: EMA; 2021 [updated 8 November 2021]. Available from: https://www.ema.europa.eu/en/documents/product-information/votrient-epar-product-information_en.pdf.

223. European Medicines Agency (EMA). Fotivda. Summary of product characteristics. Amsterdam: EMA; 2022 [updated 25 July 2022]. Available from: https://www.ema.europa.eu/en/documents/product-information/fotivda-epar-product-information_en.pdf.

224. European Medicines Agency (EMA). Sutent. Summary of product characteristics. Amsterdam: EMA; 2021 [updated 13 August 2021]. Available from: https://www.ema.europa.eu/en/documents/product-information/sutent-epar-product-information_en.pdf.

225. European Medicines Agency (EMA). Opdivo. Summary of product characteristics. Amsterdam: EMA; 2022 [updated 28 March 2023]. Available from: https://www.ema.europa.eu/en/documents/product-information/opdivo-epar-product-information_en.pdf.

226. European Medicines Agency (EMA). Kisplyx. Summary of product characteristics. Amsterdam: EMA; 2022 [updated 17 June 2022]. Available from: https://www.ema.europa.eu/documents/product-information/kisplyx-epar-product-information_en.pdf.

227. European Medicines Agency (EMA). Keytruda. Summary of product characteristics. Amsterdam: EMA; 2022 [updated 17 November 2022]. Available from: https://www.ema.europa.eu/en/documents/product-information/keytruda-epar-product-information_en.pdf.

228. National Institute for Health and Care Excellence (NICE). Lu vipivotide tetraxetan for treating PSMA-positive hormone-relapsed metastatic prostate cancer after 2 or more therapies [ID3840]. In development [GID-TA10730]. Project documents. London: NICE; 2023 [updated 21 February 2023]. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10730/documents.

229. The Uromigos. ASCO 2023 emergency podcast IPINIVO vs VEGFPD1 in kidney cancer. Podcast. Episode 249. Spotify for Podcasters; 2023 [updated 18 June 2023]. Available from: https://podcasters.spotify.com/pod/show/the-uromigos/episodes/1st-ASCO23-emergency-podcast-IPINIVO-vs-VEGFPD1-in-kidney-cancer-e25rmji.

230. Office for National Statistics (ONS). National life tables: England and Wales. 2021 [updated 23 September 2021]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandandwalesreferencetables.

231. Ambavane A, Yang S, Atkins MB, Rao S, Shah A, Regan MM, et al. Clinical and economic outcomes of treatment sequences for intermediate- to poor-risk advanced renal cell carcinoma. Immunotherapy. 2020;12(1):37-51.

232. Bensimon AG, Zhong Y, Swami U, Briggs A, Young J, Feng Y, et al. Cost-effectiveness of pembrolizumab with axitinib as first-line treatment for advanced renal cell carcinoma. Curr Med Res Opin. 2020;36(9):1507-17.

233. Haddad R, Cohen EEW, Venkatachalam M, Young K, Singh P, Shaw JW, et al. Cost-effectiveness analysis of nivolumab for the treatment of squamous cell carcinoma of the head and neck in the United States. J Med Econ. 2020;23(5):442-7.

234. National Institute for Health and Care Excellence (NICE). Larotrectinib for treating NTRK fusion-positive solid tumours. Technology appraisal guidance [TA630]. London: NICE; 2020. Available from: https://www.nice.org.uk/guidance/ta630.

235. Meng J, Lister J, Vataire AL, Casciano R, Dinet J. Cost-effectiveness comparison of cabozantinib with everolimus, axitinib, and nivolumab in the treatment of advanced renal cell carcinoma following the failure of prior therapy in England. Clinicoecon Outcomes Res. 2018;10:243-50.

236. Amdahl J, Diaz J, Sharma A, Park J, Chandiwana D, Delea TE. Cost-effectiveness of pazopanib versus sunitinib for metastatic renal cell carcinoma in the United Kingdom. PLoS One. 2017;12(6):e0175920.

237. Porta CG, Motzer R, Ejzykowicz F, Blum SI, Hamilton M, May JR, et al. 668P Matching-adjusted indirect comparison (MAIC) of health-related quality of life (HRQoL) of nivolumab plus cabozantinib (N+C) vs pembrolizumab plus axitinib (P+A) in previously untreated advanced renal cell carcinoma (aRCC). Ann Oncol. 2021;32:S692-S3.

238. Henegan JC, Bell FG. Variance of mean baseline health-related quality of life across phase III clinical trials in metastatic renal cell carcinoma. J Clin Oncol. 2022;40(6_suppl):329-.

239. Motzer RJ, Porta C, Alekseev B, Rha SY, Choueiri TK, Mendez-Vidal MJ, et al. Health-related quality-of-life (HRQoL) analysis from the phase 3 CLEAR trial of lenvatinib (LEN) plus pembrolizumab (PEMBRO) or everolimus (EVE) versus sunitinib (SUN) for patients (pts) with advanced renal cell carcinoma (aRCC). J Clin Oncol. 2021;39(15_suppl):4502-.

240. Mouillet G, Pozet A, Fritzsch J, Es-Saad I, Paget-Bailly S, Foubert A, et al. Heath related quality of life (HRQOL) assessment for patients with advanced renal cell carcinoma (mRCC) treated with tyrosine kinase inhibitor (TKI) using electronic patient reported outcome (PRO) in daily clinical practice. Ann Oncol. 2017;28:v565.

241. Cella D, Grünwald V, Escudier B, Hammers HJ, George S, Nathan P, et al. Patient-reported outcomes of patients with advanced renal cell carcinoma treated with nivolumab plus ipilimumab versus sunitinib (CheckMate 214): a randomised, phase 3 trial. Lancet Oncol. 2019;20(2):297-310.

242. Cella D, Motzer RJ, Blum SI, Ejzykowicz F, Hamilton M, Wallace J, et al. Health-related quality of life (HRQoL) in previously untreated patients with advanced renal cell carcinoma (aRCC): CheckMate 9ER updated results. J Clin Oncol. 2022;40(6_suppl):323-.

243. Bedke J, Rini BI, Plimack ER, Stus V, Gafanov R, Waddell T, et al. Health-related Quality of Life Analysis from KEYNOTE-426: Pembrolizumab plus Axitinib Versus Sunitinib for Advanced Renal Cell Carcinoma. Eur Urol. 2022;82(4):427-39.

244. Buckley HL, Collinson FJ, Ainsworth G, Poad H, Flanagan L, Katona E, et al. PRISM protocol: a randomised phase II trial of nivolumab in combination with alternatively scheduled ipilimumab in first-line treatment of patients with advanced or metastatic renal cell carcinoma. BMC Cancer. 2019;19(1):1102.

245. Swinburn P, Lloyd A, Nathan P, Choueiri TK, Cella D, Neary MP. Elicitation of health state utilities in metastatic renal cell carcinoma. Curr Med Res Opin. 2010;26(5):1091-6.

246. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708-15.

247. National Institute for Health and Care Excellence (NICE). NICE DSU Technical Support Document 12: The Use of Health State Utility Values in Decision Models. London: NICE; 2011. Available from: https://www.ncbi.nlm.nih.gov/books/NBK425824/.

248. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health. 2010;13(5):509-18.

249. Alava MH, Pudney S, Wailoo A. Estimating EQ-5D by age and sex for the UK. Report by the Decision Support Unit. Sheffield: School of Health and Related Research (ScHARR); 2022. Available from: https://www.sheffield.ac.uk/sites/default/files/2022-02/DSU%20Age%20based%20utility%20-%20Final%20for%20website.pdf.

250. Hansen RN, Hackshaw MD, Nagar SP, Arondekar B, Deen KC, Sullivan SD, et al. Health care costs among renal cancer patients using pazopanib and sunitinib. J Manag Care Spec Pharm. 2015;21(1):37-44, a-d.

251. Reference costs. 2011-12. London: Department of Health; 2012. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213060/2011-12-reference-costs-publication.pdf.

252. Office for National Statistics (ONS). CPI Weights 06: Health. Consumer Price Indices for Health 1988–2014. 2023 [updated 22 March 2023]. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/chzw/mm23.

253. Reference Costs 2014–15. London: Department of Health; 2015. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/477919/2014-15_Reference_costs_publication.pdf.

254. Curtis L, Burns A. Unit Costs of Health and Social Care 2015. Canterbury: Personal Social Services Research Unit, University of Kent; 2015. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2015/.

255. ClinicalTrials.gov. A Study of Cabozantinib (XL184) vs Everolimus in Subjects With Metastatic Renal Cell Carcinoma (METEOR). ClinicalTrials.gov Identifier: NCT01865747. 2017 [updated 27 April 2021]. Available from: https://clinicaltrials.gov/ct2/show/NCT01865747.

256. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. London: Nuffield Trust; 2014. Available from: https://www.nuffieldtrust.org.uk/files/2017-01/end-of-life-care-web-final.pdf.

257. Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T, et al. Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. Health Technol Assess. 2010;14(2):1-184, iii-iv.

258. Edwards SJ, Wakefield V, Cain P, Karner C, Kew K, Bacelar M, et al. Axitinib, cabozantinib, everolimus, nivolumab, sunitinib and best supportive care in previously treated renal cell carcinoma: a systematic review and economic evaluation. Health Technol Assess. 2018;22(6):1-278.

259. Jones KC, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A, et al. Unit Costs of Health and Social Care 2022 Manual. Technical report. Canterbury: Personal Social Services Research Unit, University of Kent; 2023. Available from: https://kar.kent.ac.uk/100519/1/Unit_Costs_of_Health_and_Social_Care_2022%20%287%29.pdf.

260. Curtis L, Burns A. Unit Costs of Health and Social Care 2017. Canterbury: Personal Social Services Research Unit, University of Kent; 2017. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2017/.

261. British National Formulary (BNF). London: BMJ & Royal Pharmaceutical Society; 2023 [updated 31 March 2023]. Available from: https://bnf.nice.org.uk (accessed 31 March 2023).

262. Drugs and pharmaceutical electronic market information tool (eMIT). London: Department of Health and Social Care; 2023 [updated 22 March 2023]. Available from: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit.

263. NHS England. National Schedule of NHS Costs 2021/22. V3. 2023 [updated 19 May 2023]. Available from: https://www.england.nhs.uk/publication/2021-22-national-cost-collection-data-publication/ (accessed 25 June 2023).

264. Spencer K, Parrish R, Barton R, Henry A. Palliative radiotherapy. BMJ. 2018;360:k821.

265. Iuliani M, Simonetti S, Pantano F, Ribelli G, Di Martino A, Denaro V, et al. Antitumor Effect of Cabozantinib in Bone Metastatic Models of Renal Cell Carcinoma. Biology (Basel). 2021;10(8).

266. Ratta R, Verzoni E, Mennitto A, Pantano F, Martinetti A, Raimondi A, et al. Effects of cabozantinib on bone turnover markers in real-world metastatic renal cell carcinoma. Tumori. 2021;107(6):542-9.

267. Bersanelli M, Buti S, Ghidini A, Tiseo M, Petrelli F. A metanalysis on cabozantinib and bone metastases: true story or commercial gimmick? Anticancer Drugs. 2020;31(3):211-5.

Appendices

Appendix A: Literature search strategies

Appendix B: PRISMA diagrams for clinical review

Appendix C: Excluded studies

Appendix D: Data extraction grids and quality assessment

Appendix E: Network meta-analysis additional materials

Appendix F: Targeted review of surrogacy in advanced RCC

Appendix G: Company estimation of health state utilities

Appendix H: Additional HRQoL studies

Appendix I: Consistency with previous technology appraisals

Appendix J: Structured expert elicitation materials

Appendix K: Supplementary survival analysis information

Appendix L: Real-world evidence supplementary information

Appendix M: Clinician feedback to the EAG

Appendix N: Subsequent treatment proportions

Appendix O: Adverse event rates and costs (trial scenario)